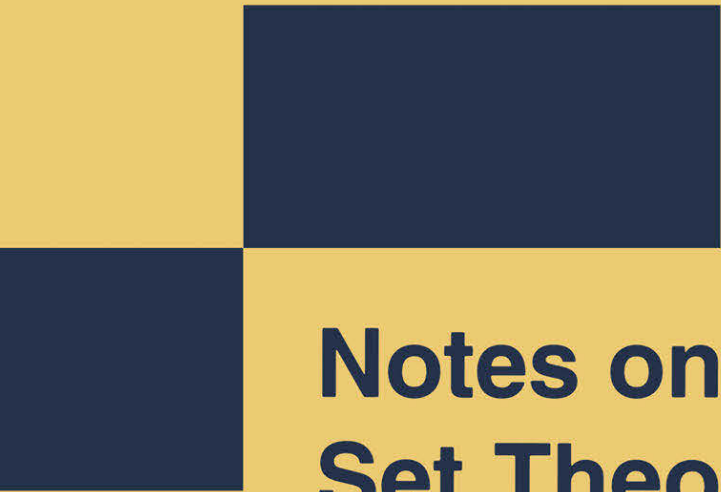


Yiannis N. Moschovakis



Notes on Set Theory



Springer

Undergraduate Texts in Mathematics

Editors

J.H. Ewing
F.W. Gehring
P.R. Halmos

Undergraduate Texts in Mathematics

Apostol: Introduction to Analytic Number Theory. Second edition.

Armstrong: Groups and Symmetry.

Armstrong: Basic Topology.

Bak/Newman: Complex Analysis.

Banchhoff/Wermer: Linear Algebra Through Geometry. Second edition.

Brémaud: An Introduction to Probabilistic Modeling.

Bressoud: Factorization and Primality Testing.

Bressoud: Second Year Calculus.

Readings in Mathematics.

Brickman: Mathematical Introduction to Linear Programming and Game Theory.

Cederberg: A Course in Modern Geometries.

Childs: A Concrete Introduction to Higher Algebra.

Chung: Elementary Probability Theory with Stochastic Processes. Third edition.

Cox/Little/O'Shea: Ideals, Varieties, and Algorithms.

Curtis: Linear Algebra: An Introductory Approach. Fourth edition.

Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.

Dixmier: General Topology.

Driver: Why Math?

Ebbinghaus/Flum/Thomas: Mathematical Logic.

Edgar: Measure, Topology, and Fractal Geometry.

Fischer: Intermediate Real Analysis.

Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.

Fleming: Functions of Several Variables. Second edition.

Foulds: Optimization Techniques: An Introduction.

Foulds: Combinatorial Optimization for Undergraduates.

Franklin: Methods of Mathematical Economics.

Halmos: Finite-Dimensional Vector Spaces. Second edition.

Halmos: Naive Set Theory.

Hämmerlin/Hoffmann: Numerical Mathematics.

Readings in Mathematics.

Iooss/Joseph: Elementary Stability and Bifurcation Theory. Second edition.

James: Topological and Uniform Spaces.

Jänich: Topology.

Klambauer: Aspects of Calculus.

Kinsey: Topology of Surfaces.

Lang: A First Course in Calculus. Fifth edition.

Lang: Calculus of Several Variables. Third edition.

Lang: Introduction to Linear Algebra. Second edition.

Lang: Linear Algebra. Third edition.

Lang: Undergraduate Algebra. Second edition.

Lang: Undergraduate Analysis.

Lax/Burstein/Lax: Calculus with Applications and Computing. Volume 1.

LeCuyer: College Mathematics with APL.

(continued after index)

Yiannis N. Moschovakis

Notes on Set Theory

With 36 Illustrations



Springer

Yiannis N. Moschovakis
Department of Mathematics
University of California
Los Angeles, CA 90024 USA

Editorial Board:

John H. Ewing
Department of Mathematics
Indiana University
Bloomington, IN 47405 USA

F.W. Gehring
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109 USA

Paul R. Halmos
Department of Mathematics
Santa Clara University
Santa Clara, CA 95053 USA

Mathematics Subject Classifications (1991): 04-01

Library of Congress Cataloging-in-Publication Data
Moschovakis, Yiannis N.

Notes on set theory / Yiannis N. Moschovakis.
p. cm. — (Undergraduate texts in mathematics)
Includes bibliographical references and index.

1. Set theory. I. Title. II. Series.
QA248.M665 1994
511.3'22—dc20 93-35825

© 1994 Springer Science+Business Media New York Greek edition published as *Notes in Set Theory* by Nefeli Publications, Athens, 1993.

Originally published by Springer-Verlag New York, Inc in 1994.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher Springer Science+Business Media, LLC except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Dedicated to the memory of Nikos Kritikos

PREFACE

What this book is about. The *theory of sets* is a vibrant, exciting mathematical theory, with its own basic notions, fundamental results and deep open problems, and with significant applications to other mathematical theories. At the same time, *axiomatic set theory* is often viewed as a *foundation of mathematics*: it is alleged that all mathematical objects are sets, and their properties can be derived from the relatively few and elegant axioms about sets. Nothing so simple-minded can be quite true, but there is little doubt that in standard, current mathematical practice, “making a notion precise” is essentially synonymous with “defining it in set theory.” Set theory is the official language of mathematics, just as mathematics is the official language of science.

Like most authors of elementary, introductory books about sets, I have tried to do justice to both aspects of the subject.

From straight set theory, these Notes cover the basic facts about “abstract sets,” including the Axiom of Choice, transfinite recursion, and cardinal and ordinal numbers. Somewhat less common is the inclusion of a chapter on “pointsets” which focuses on results of interest to analysts and introduces the reader to the Continuum Problem, central to set theory from the very beginning. There is also some novelty in the approach to cardinal numbers, which are brought in very early (following Cantor, but somewhat deviously), so that the basic formulas of cardinal arithmetic can be taught as quickly as possible. Appendix A gives a more detailed “construction” of the real numbers than is common nowadays, which in addition claims some novelty of approach and detail. Appendix B is a somewhat eccentric, mathematical introduction to the study of *natural models* of various set theoretic principles, including Aczel’s Antifoundation. It assumes no knowledge of logic, but should drive the serious reader to study it.

About set theory as a foundation of mathematics, there are two aspects of these Notes which are somewhat uncommon. First, I have taken seriously this business about “everything being a set” (which of course it is not) and have tried to make sense of it in terms of the notion of *faithful representation* of mathematical objects by *structured sets*. An old idea, but perhaps this is the first textbook which takes it seriously, tries to explain it, and applies

it consistently. Those who favor category theory will recognize some of its basic notions in places, shamelessly folded into a traditional set theoretical approach to the foundations where categories are never mentioned. Second, *computation theory* is viewed as part of the mathematics “to be founded” and the relevant set theoretic results have been included, along with several examples. The ambition was to explain what every young mathematician or theoretical computer scientist needs to know about sets.

The book includes several historical remarks and quotations which in some places give it an undeserved scholarly gloss. All the quotations (and most of the comments) are from papers reprinted in the following two marvelous and easily accessible source books, which should be perused by all students of set theory:

Georg Cantor, *Contributions to the founding of the theory of transfinite numbers*, translated and with an Introduction by Philip E. B. Jourdain, Dover Publications, New York.

Jean van Heijenoort, *From Frege to Gödel*, Harvard University Press, Cambridge, 1967.

How to use it. About half of this book can be covered in a Quarter (ten weeks), somewhat more in a longer Semester. Chapters 1 - 6 cover the beginnings of the subject and they are written in a leisurely manner, so that the serious student can read through them alone, with little help. The trick to using the Notes successfully in a class is to cover these beginnings very quickly: skip the introductory Chapter 1, which mostly sets notation; spend about a week on Chapter 2, which explains Cantor’s basic ideas; and then proceed with all deliberate speed through Chapters 3 - 6, so that the theory of well ordered sets in Chapter 7 can be reached no later than the sixth week, preferably the fifth. Beginning with Chapter 7, the results are harder and the presentation is more compact. How much of the “real” set theory in Chapters 7 - 12 can be covered depends, of course, on the students, the length of the course, and what is passed over. If the class is populated by future computer scientists, for example, then Chapter 6 on Fixed Points should be covered in full, with its problems, but Chapter 10 on Baire Space might be omitted, sad as that sounds. For budding young analysts, at the other extreme, Chapter 6 can be cut off after 6.26 (and this too is sad), but at least part of Chapter 10 should be attempted. Additional material which can be left out, if time is short, includes the detailed development of addition and multiplication on the natural numbers in Chapter 5, and some of the less central applications of the Axiom of Choice in Chapter 9. The Appendices are quite unlikely to be taught in a course (I devote just one lecture to explain the idea of the construction of the reals in Appendix A), though I would like to think that they might be suitable for undergraduate Honors Seminars, or individual reading courses.

Since elementary courses in Set Theory are not offered regularly and

they are seldom long enough to cover all the basics, I have tried to make these Notes accessible to the serious student who is studying the subject on his own. There are numerous, simple Exercises strewn throughout the text, which test understanding of new notions immediately after they are introduced. In class I present about half of them, as examples, and I assign some of the rest for easy homework. The Problems at the end of each chapter vary widely in difficulty, some of them covering additional material. The hardest problems are marked with an asterisk (*).

Acknowledgments. I am grateful to the Mathematics Department of the University of Athens for the opportunity to teach there in Fall 1990, when I wrote the first draft of these Notes, and especially to Prof. A. Tsarpalias, who usually teaches that Set Theory course and used a second draft in Fall 1991; and to Dimitra Kitsiou and Stratos Paschos for struggling with PCs and laser printers at the Athens Polytechnic in 1990 to produce the first “hard copy” version. I am grateful to my friends and colleagues at UCLA and Caltech (hotbeds of activity in set theory) from whom I have absorbed what I know of the subject, over many years of interaction. I am especially grateful to my wife Joan Moschovakis and my student Darren Kessner for reading large parts of the preliminary edition, doing the problems and discovering a host of errors; and to Larry Moss who taught out of the preliminary edition in the Spring Term of 1993, found the remaining host of errors and wrote out solutions to many of the problems.

The book was written more-or-less simultaneously in Greek and English, by the magic of bilingual $\text{\LaTeX}^1$ and in true reflection of my life. I have dedicated it to Prof. Nikos Kritikos (a student of Caratheodory), in fond memory of many unforgettable hours he spent with me back in 1973, patiently teaching me how to speak and write mathematics in my native tongue, but also much about the love of science and the nature of scholarship. In this connection, I am also greatly indebted to Takis Koufopoulos, who read critically the preliminary Greek version, corrected a host of errors and made numerous suggestions which (I believe) improved substantially the language of the final Greek draft.

Palaion Phaliron, Greece

November 1993

¹Greek mathematicians owe a substantial debt of gratitude to Silvio Levy, whose lovely Greek fonts made it possible to use $\text{\TeX}$ and $\text{\LaTeX}$ in Greek scientific publishing. Those interested in the package of public domain programs I have used to typeset the Greek edition should contact me by mail, or (preferably) electronically at ynm@math.ucla.edu.

CONTENTS

1. Introduction	1
Problems for Chapter 1	5
2. Equinumerosity	7
Countable unions of countable sets	9
The reals are uncountable	11
$A <_c \mathcal{P}(A)$	15
Schröder-Bernstein Theorem	16
Problems for Chapter 2	18
3. Paradoxes and axioms	19
The Russell paradox	21
Axioms (I) - (VI)	24
Axioms for definite conditions and operations	27
Classes	28
Problems for Chapter 3	31
4. Are sets all there is?	33
Ordered pairs	35
Disjoint union	36
Relations	37
Equivalence relations	38
Functions	39
Cardinal numbers	43
Structured sets	45
Problems for Chapter 4	46

5. The natural numbers	53
Existence of the Natural Numbers	54
Uniqueness of the Natural Numbers	54
Recursion Theorem	55
Addition and multiplication	59
Pigeonhole Principle	64
Strings	67
The continuum	69
Problems for Chapter 5	69
6. Fixed points	73
Posets	73
Partial functions	76
Inductive posets	77
Continuous Least Fixed Point Theorem	79
About topology	81
Graphs	85
Problems for Chapter 6	86
Streams	87
Scott topology	91
Directed-complete posets	91
7. Well ordered sets	93
Transfinite induction	98
Transfinite recursion	100
Iteration Lemma	100
Comparability of well ordered sets	104
Wellfoundedness of $\leq_o$	105
Hartogs' Theorem	106
Fixed Point Theorem	108
Least Fixed Point Theorem	108
Problems for Chapter 7	110
8. Choices	117
Axiom of Choice	117
Equivalents of AC	120
Countable Principle of Choice, AC_N	122
Axiom (VI) of Dependent Choices, DC	122
The axiomatic theories ZDC , ZAC	125
Consistency and independence results	126
Problems for Chapter 8	127

9. Choice's consequences	131
Trees	132
König's Lemma	133
Fan Theorem	134
Wellfoundedness of $\leq_c$	134
Best wellorderings	135
Absorption laws	138
König's Theorem	140
Cofinality, regular cardinals	141
Problems for Chapter 9	142
 10. Baire space	 147
Cardinality of perfect pointsets	150
Cantor-Bendixson Theorem	151
Property P	152
Analytic pointsets	153
Perfect Set Theorem	157
Borel sets	160
Counterexample to the general property P	162
Consistency and independence results	164
Problems for Chapter 10	165
Borel isomorphisms	166
 11. Replacement and other axioms	 169
Replacement Axiom (VIII)	170
The axiomatic theories ZFDC , ZFAC	170
Grounded Recursion Theorem	172
Transitive classes	174
Basic Closure Lemma	175
Hereditarily finite sets	176
Zermelo universes	177
The least Zermelo universe	179
Grounded sets	180
Principle of Foundation	180
The axiomatic theory Zermelo-Fraenkel, ZFC	181
Z-F universes	183
von Neumann's class $\mathcal{V}$	183
Mostowski Collapsing Lemma	183
Consistency and independence results	184
Problems for Chapter 11	185

12. Ordinal numbers	189
Characterization of the ordinal assignment	193
Characterization of the ordinals	194
Ordinal recursion	197
Ordinal addition, multiplication	197
von Neumann cardinals	198
The operation $\aleph_\alpha$	200
The cumulative rank hierarchy	201
Problems for Chapter 12	203
The operation $\beth_\alpha$	205
Strongly inaccessible cardinals	206
Frege cardinals	206
Quotients of equivalence conditions	207
A. The real numbers	209
Congruences	209
Fields	211
Ordered fields	212
Uniqueness of the rationals	214
Existence of the rationals	215
Countable, dense, linear orderings	219
The archimedean property	221
Nested interval property	226
Dedekind cuts	229
Existence of the real numbers	231
Uniqueness of the real numbers	234
Problems for Appendix A	236
B. Axioms and universes	239
Set universes	242
Propositions and relativizations	243
Rieger universes	248
Rieger's Theorem	248
Antifoundation Principle, AFA	254
Bisimulations	255
The antifounded universe	259
Aczel's Theorem	259
Problems for Appendix B	262
Index	267

Chapter 1

INTRODUCTION

Mathematicians have always used sets, e.g. the ancient Greek geometers defined a circle as the set of points at a fixed distance r from a fixed point C , its center. But the systematic study of sets began only at the end of the 19th century with the work of the great German mathematician Georg Cantor, who created a rigorous theory of the concept of *completed infinite* by which we can compare infinite sets as to size. For example, let

$N = \{0, 1, \dots\}$ = the set of natural numbers,

Q = the set of rational numbers (fractions),

$\mathcal{R}$ = the points of a straight line,

where we also identify $\mathcal{R}$ with the set of real numbers, each point associated with its (positive or negative) coordinate with respect to a fixed origin and direction. Cantor asked if these three sets “have the same (infinite) number of elements,” or if one of them is “more numerous” than the others. Before we make precise and answer this question in the next chapter, we review here some basic, well-known facts about sets and functions, primarily to explain the notation we will be using.

What are sets, anyway? The question is like “what are points,” which Euclid answered with

a point is that which has no parts.

This is not a rigorous mathematical definition, a reduction of the concept of “point” to other concepts which we already understand, but just an intuitive description which suggests that a point is some thing which has no extension in space. Like that of point, the concept of set is fundamental and cannot be reduced to other, simpler concepts. Cantor described it as follows:

By a set we are to understand any collection into a whole of definite and separate objects of our intuition or our thought.

Vague as it is, this description implies two basic properties of sets.

1. Every set A has **elements** or **members**. We write

$x \in A \iff$ the object x is a member of (or belongs to) A .

2. A set is determined by its members, i.e. if A, B are sets, then¹

$$A = B \iff (\forall x)[x \in A \iff x \in B].$$

This last is the **extensionality property**. For example, the set of students in this class will not change if we all switch places, lie down or move to another classroom; this set is completely determined by *who we are*, not our posture or the places where we happen to be.

Somewhat peculiar is the **empty set** $\emptyset$ which has no members. The extensionality property implies that *there is only one empty set*.

If A and B are sets, we write

$$A \subseteq B \iff (\forall x)[x \in A \implies x \in B],$$

and if $A \subseteq B$, we call A a **subset** of B , so that for every B ,

$$\emptyset \subseteq B, \quad B \subseteq B.$$

A **proper subset** of B is a subset distinct from B ,

$$A \subsetneq B \iff [A \subseteq B \ \& \ A \neq B].$$

From the extensionality property it follows that for any two sets A, B ,

$$A = B \iff A \subseteq B \ \& \ B \subseteq A.$$

We have already used several different notations to define specific sets and we need still more, e.g.

$$A = \{a_1, a_2, \dots, a_n\}$$

is the (finite) set with members the objects $a_1, a_2, \dots, a_n$. If P is a condition which specifies some property of an arbitrary object x , then

$$A = \{x \mid P(x)\}$$

¹We will use systematically, as abbreviations, the logical symbols

$\&$: and, $\vee$: or, $\neg$: not, $\implies$: implies, $\iff$: if and only if,

$\forall$: for all, $\exists$: there exists, $\exists!$: there exists exactly one.

is the set of all objects which satisfy the condition P , so that for all x ,

$$x \in A \iff P(x).$$

For any two sets A, B ,

$$\begin{aligned} A \cup B &= \{x \mid x \in A \vee x \in B\} && \text{(the union of } A, B), \\ A \cap B &= \{x \mid x \in A \ \& \ x \in B\} && \text{(the intersection of } A, B), \\ A \setminus B &= \{x \mid x \in A \ \& \ x \notin B\} && \text{(the difference of } A, B). \end{aligned}$$

The union and the intersection of a sequence of sets are defined in the same way,

$$\begin{aligned} \bigcup_{n=0}^{\infty} A_n &= \{x \mid (\exists n \in N)[x \in A_n]\}, \\ \bigcap_{n=0}^{\infty} A_n &= \{x \mid (\forall n \in N)[x \in A_n]\}. \end{aligned}$$

We will use the notation

$$f : X \rightarrow Y$$

to indicate that f is a **function** which associates with each member x of the set X some member $f(x)$ of Y . Functions are also called **mappings**, **operations**, **transformations** and many other things. Sometimes it is convenient to use the abbreviated notation $(x \mapsto f(x))$ which makes it possible to talk about a function without officially naming it. For example,

$$(x \mapsto x^2 + 1)$$

is the function on the real numbers which assigns to each real its square increased by 1; if we call it f , then it is defined by the formula

$$f(x) = x^2 + 1 \quad (x \in \mathcal{R})$$

so that $f(0) = 1$, $f(2) = 5$, etc. But we can say “all the values of $(x \mapsto x^2 + 1)$ are positive reals” without necessarily fixing a name for it, like f .

In connection with functions we will also use the notations

$$\begin{aligned} f : X \rightarrowtail Y &\iff_{\text{df}} f \text{ is an } \mathbf{injection} \text{ (one-to-one)} \\ &\iff (\forall x, x' \in X)[f(x) = f(x') \implies x = x'], \\ f : X \twoheadrightarrow Y &\iff_{\text{df}} f \text{ is a } \mathbf{surjection} \text{ (onto)} \\ &\iff (\forall y \in Y)(\exists x \in X)[f(x) = y], \\ f : X \xrightarrow{\sim} Y &\iff_{\text{df}} f \text{ is a } \mathbf{bijection} \text{ or } \mathbf{correspondence} \\ &\iff (\forall y \in Y)(\exists! x \in X)[f(x) = y]. \end{aligned}$$

For every $f : X \rightarrow Y$ and $A \subseteq X$, the set

$$f[A] = \{f(x) \mid x \in A\}$$

is the **image** of A under f , and if $B \subseteq Y$, then

$$f^{-1}[B] = \{x \in X \mid f(x) \in B\}$$

is the **pre-image** of B by f .

If f is a bijection, then we can define the **inverse function** $f^{-1} : Y \rightarrow X$ by the condition

$$f^{-1}(y) = x \iff f(x) = y,$$

and then the inverse image $f^{-1}[B]$ (as we defined it above) is precisely the image of B under f^{-1} .

The **composition**

$$h =_{\text{df}} gf : X \rightarrow Z$$

of two functions

$$f : X \rightarrow Y, \quad g : Y \rightarrow Z$$

is defined by

$$h(x) = g(f(x)) \quad (x \in X).$$

It is easy to prove many basic properties of sets and functions using only these definitions and the extensionality property. For example, if $f : X \rightarrow Y$ and $A, B \subseteq X$, then

$$f[A \cup B] = f[A] \cup f[B].$$

To prove this, we verify that an arbitrary $y \in Y$ belongs to $f[A \cup B]$ if and only if it is a member of $f[A] \cup f[B]$:

$$\begin{aligned} y \in f[A \cup B] &\iff (\exists x)[x \in A \cup B \ \& \ y = f(x)] \\ &\iff (\exists x)[(x \in A \vee x \in B) \ \& \ y = f(x)] \\ &\iff (\exists x)[x \in A \ \& \ y = f(x)] \vee (\exists x)[x \in B \ \& \ y = f(x)] \\ &\iff y \in f[A] \vee y \in f[B] \\ &\iff y \in f[A] \cup f[B]. \end{aligned}$$

In some cases, the logic of the argument gets a bit complex and it is easier to prove an identity $U = V$ by verifying separately the implications $x \in U \implies x \in V$ and $x \in V \implies x \in U$.

Problems

x1.1. For any three sets A, B, C ,

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C), \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \setminus (A \cap B) &= A \setminus B. \end{aligned}$$

x1.2. (De Morgan's laws.) For any three sets A, B, C ,

$$\begin{aligned} C \setminus (A \cup B) &= (C \setminus A) \cap (C \setminus B), \\ C \setminus (A \cap B) &= (C \setminus A) \cup (C \setminus B). \end{aligned}$$

x1.3. For every injection $f : X \rightarrow Y$, and all $A, B \subseteq X$,

$$\begin{aligned} f[A \cap B] &= f[A] \cap f[B], \\ f[A \setminus B] &= f[A] \setminus f[B]. \end{aligned}$$

Show also that these identities do not always hold if f is not an injection.

x1.4. For every $f : X \rightarrow Y$, and all $A, B \subseteq Y$,

$$\begin{aligned} f^{-1}[A \cup B] &= f^{-1}[A] \cup f^{-1}[B], \\ f^{-1}[A \cap B] &= f^{-1}[A] \cap f^{-1}[B]. \end{aligned}$$

x1.5. For every $f : X \rightarrow Y$ and every sequence of sets $B_n \subseteq Y$,

$$\begin{aligned} f^{-1}\left[\bigcup_{n=0}^{\infty} B_n\right] &= \bigcup_{n=0}^{\infty} f^{-1}[B_n], \\ f^{-1}\left[\bigcap_{n=0}^{\infty} B_n\right] &= \bigcap_{n=0}^{\infty} f^{-1}[B_n], \\ f\left[\bigcup_{n=0}^{\infty} A_n\right] &= \bigcup_{n=0}^{\infty} f[A_n]. \end{aligned}$$

x1.6. For every injection $f : X \rightarrow Y$ and every sequence of sets $A_n \subseteq X$,

$$f\left[\bigcap_{n=0}^{\infty} A_n\right] = \bigcap_{n=0}^{\infty} f[A_n].$$

x1.7. The composition of injections is an injection, the composition of surjections is a surjection, and hence the composition of bijections is a bijection.

Chapter 2

EQUINUMEROSITY

After these preliminaries, we can formulate the fundamental definitions of Cantor about the size or cardinality of sets.

2.1. Definition. *Two sets A, B are **equinumerous** or **equal in cardinality** if there exists a (one-to-one) correspondence between their elements, in symbols*

$$A =_c B \iff_{\text{df}} (\exists f)[f : A \rightarrow B].$$

This definition of equinumerosity stems from our intuitions about finite sets, e.g. we can be sure that a shoe store offers for sale the same number of left and right shoes without knowing exactly what that number is: the correspondence of each left shoe with the right shoe in the same pair establishes the equinumerosity of these two sets. The radical element in Cantor's definition is the proposal to accept the existence of such a correspondence as the characteristic property of equinumerosity for all sets, despite the fact that its application to infinite sets leads to conclusions which had been viewed as counterintuitive. A finite set, for example, cannot be equinumerous with one of its proper subsets, while the set of natural numbers N is equinumerous with $N \setminus \{0\}$ via the correspondence $(x \mapsto x + 1)$,

$$\{0, 1, 2, \dots\} =_c \{1, 2, 3, \dots\}.$$

In the real numbers, also,

$$(0, 1) =_c (0, 2)$$

via the correspondence $(x \mapsto 2x)$, where as usual, for any two reals $\alpha < \beta$

$$(\alpha, \beta) = \{x \in \mathcal{R} \mid \alpha < x < \beta\}.$$

(We will use the analogous notation for the closed and half-closed intervals $[\alpha, \beta]$, $[\alpha, \beta)$, etc.)

2.2. Proposition. *For all sets A, B, C ,*

$$\begin{aligned} A &=_c A, \\ A =_c B &\implies B =_c A, \\ A =_c B \ \& \ B =_c C &\implies A =_c C. \end{aligned}$$

$f : \{i \mid i < n\} \rightarrow A$ with some $n > 0$, and we can set

$$\pi(i) = \begin{cases} f(i), & \text{if } i < n, \\ f(0), & \text{if } i \geq n. \end{cases}$$

To prove the converse, suppose A is not finite and it has an enumeration $\pi : N \rightarrow A$. We must find another enumeration $f : N \rightarrow A$ which is *without repetitions*, so that it is in fact a bijection of N with A , and hence A is countably infinite. The proof is suggested by Figure 2.1: we simply delete the repetitions from the given enumeration π of A . To get a precise definition of f by recursion, notice that because A is not finite, for every finite sequence $a_0, \dots, a_n$ of members of A there exists some m such that $\pi(m) \notin \{a_0, \dots, a_n\}$, and set

$$\begin{aligned} f(0) &= \pi(0), \\ m_n &= \text{the least } m > n \text{ such that } \pi(m) \notin \{f(0), \dots, f(n)\}, \\ f(n+1) &= \pi(m_n). \end{aligned}$$

It is obvious that f is an injection, so it is enough to verify that every $x \in A$ is a value of f . This is clearly true for $\pi(0) = f(0)$. If $x = \pi(n+1)$ for some n and $x \in \{f(0), \dots, f(n)\}$, then $x = f(i)$ for some $i \leq n$, and if $x \notin \{f(0), \dots, f(n)\}$, then $m_n = n+1$ and $f(n+1) = \pi(m_n) = x$ by the definition. $\dashv$

2.8. Exercise. If A is countable and there exists an injection $f : B \rightarrow A$, then B is also countable; in particular, every subset of a countable set is countable.

2.9. Exercise. If A is countable and there exists a surjection $f : A \rightarrow B$, then B is also countable.

The next, simple theorem is one of the most basic results of set theory.

2.10. Theorem. (Cantor) For each sequence $A_0, A_1, \dots$ of countable sets, the union

$$A = \bigcup_{n=0}^{\infty} A_n = A_0 \cup A_1 \cup \dots$$

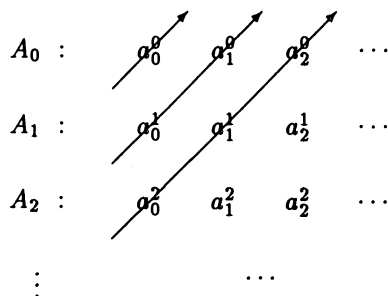
is also a countable set.

Proof. It is enough (why?) to prove the theorem in the special case where none of the A_n is empty, in which case we can find for each A_n an enumeration $\pi^n : N \rightarrow A_n$. If we let

$$a_i^n = \pi^n(i)$$

to simplify the notation, then for each n

$$A_n = \{a_0^n, a_1^n, \dots\},$$

**Figure 2.2.** Cantor's first diagonal method.

and we can construct from these enumerations a table of elements which lists all the members of the union A . The arrows in Figure 2.2 show how to enumerate the union:

$$A = \{a_0^0, a_0^1, a_1^0, a_0^2, a_1^1, \dots\}. \quad \dashv$$

2.11. Corollary. *The set of rational (positive and negative) integers*

$$\mathbb{Z} = \{\dots - 2, -1, 0, 1, 2, \dots\}$$

is countable.

Proof. $\mathbb{Z} = \mathbb{N} \cup \{-1, -2, \dots\}$ and the set of negative integers is countable via the correspondence $(x \mapsto -(x+1))$. $\dashv$

2.12. Corollary. *The set $\mathbb{Q}$ of rational numbers is countable.*

Proof. The set $\mathbb{Q}^+$ of ≥ 0 rationals is countable because

$$\mathbb{Q}^+ = \bigcup_{n=1}^{\infty} \left\{ \frac{m}{n} \mid m \in \mathbb{N} \right\}$$

and each $\{\frac{m}{n} \mid m \in \mathbb{N}\}$ is countable with the enumeration $(m \mapsto \frac{m}{n})$. The set $\mathbb{Q}^-$ of < 0 rationals is countable by the same method, and then the union $\mathbb{Q}^+ \cup \mathbb{Q}^-$ is countable by the theorem. $\dashv$

This corollary was Cantor's first significant result in the program of classification of infinite sets by their size, and it was considered somewhat "paradoxical" because $\mathbb{Q}$ appears to be so much larger than $\mathbb{N}$. Immediately afterwards, Cantor showed the existence of uncountable sets.

2.13. Theorem. (Cantor) *The set of infinite, binary sequences*

$$\Delta = \{(a_0, a_1, \dots) \mid (\forall i)[a_i = 0 \vee a_i = 1]\}$$

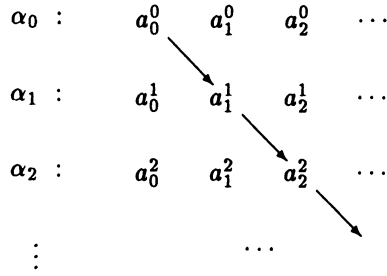


Figure 2.3. Cantor's second diagonal method.

is uncountable.

Proof. Suppose (towards a contradiction)¹ that Δ is countable, so that for some enumeration

$$\Delta = \{\alpha_0, \alpha_1, \dots\},$$

where for each n ,

$$\alpha_n = (a_0^n, a_1^n, \dots)$$

is a sequence of 0's and 1's. We construct a table with these sequences as before, and then we define the sequence β by interchanging 0 and 1 in the “diagonal” sequence $a_0^0, a_1^1, \dots$:

$$\beta(n) = 1 - a_n^n.$$

It is obvious that for each α_n , $\beta \neq \alpha_n$, since

$$\beta(n) = 1 - \alpha_n(n) \neq \alpha_n(n),$$

so that the sequence $\alpha_0, \alpha_1, \dots$ does not enumerate the entire Δ , contrary to our hypothesis. $\neg$

2.14. Corollary. (Cantor) *The set $\mathcal{R}$ of real numbers is uncountable.*

Proof. We define first a sequence of sets $\mathcal{C}_0, \mathcal{C}_1, \dots$, of real numbers which satisfy the following conditions:

1. $\mathcal{C}_0 = [0, 1]$.

¹To prove a proposition θ by the method of *reduction to a contradiction*, we assume its negation $\neg\theta$ and derive from that assumption something which violates known facts, a contradiction, something *absurd*: we conclude that θ cannot be false, so it must be true. Typically we will begin such arguments with the codephrase *towards a contradiction*, which alerts the reader that the supposition which follows is the negation of what we intend to prove.

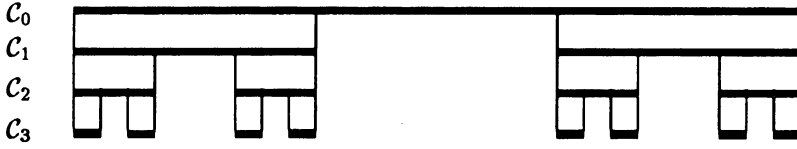


Figure 2.4. The first four stages of the Cantor set construction.

2. Each C_n is a union of 2^n closed intervals and

$$C_0 \supseteq C_1 \supseteq \cdots C_n \supseteq C_{n+1} \supseteq \cdots$$

3. C_{n+1} is constructed by removing the (open) middle third of each interval in C_n , i.e. replacing each $[a, b]$ in C_n by the two closed intervals

$$\begin{aligned} L[a, b] &= [a, a + \tfrac{1}{3}(b - a)], \\ R[a, b] &= [a + \tfrac{2}{3}(b - a), b]. \end{aligned}$$

With each binary sequence $\delta \in \Delta$ we associate now a sequence of closed intervals,

$$F_0^\delta, F_1^\delta, \dots,$$

by the following recursion:

$$\begin{aligned} F_0^\delta &= C_0 = [0, 1], \\ F_{n+1}^\delta &= \begin{cases} LF_n^\delta, & \text{if } \delta(n) = 0, \\ RF_n^\delta, & \text{if } \delta(n) = 1. \end{cases} \end{aligned}$$

By induction, for each n , F_n^δ is one of the closed intervals of C_n of length 3^{-n} and obviously

$$F_0^\delta \supseteq F_1^\delta \supseteq \cdots,$$

so by the fundamental **completeness property** of the real numbers the intersection of this sequence is not empty; in fact, it contains exactly one real number, call it

$$f(\delta) = \text{the unique element in the intersection } \bigcap_{n=0}^{\infty} F_n^\delta.$$

The function f maps the uncountable set Δ into the set

$$C = \bigcap_{n=0}^{\infty} C_n,$$

the so-called **Cantor set**, so to complete the proof it is enough to verify that f is one-to-one. But if n is the least number for which $\delta(n) \neq \varepsilon(n)$ and (for example) $\delta(n) = 0$, we have $F_n^\delta = F_n^\varepsilon$ from the choice of n , $f(\delta) \in F_{n+1}^\delta = LF_n^\delta$, $f(\varepsilon) \in F_{n+1}^\varepsilon = RF_n^\delta$, and $LF_n^\delta \cap RF_n^\delta = \emptyset$, so that indeed f is an injection. $\dashv$

The basic mathematical ingredient of this proof is the appeal to the completeness property of the real numbers, which we will study carefully in Appendix A. Some use of a special property of the reals is necessary: the rest of Cantor's construction relies solely on arithmetical properties of numbers which are also true of the rationals, so if we could avoid using completeness we would also prove that Q is uncountable, contradicting 2.12.

The fundamental importance of this theorem was instantly apparent, the more so because Cantor used it immediately in a significant application to the theory of algebraic numbers. Before we prove this corollary we need some definitions and lemmas.

2.15. Definition. For any two sets² A, B , the set of ordered pairs of members of A and members of B is denoted by

$$A \times B = \{(x, y) \mid x \in A \text{ \& } y \in B\}.$$

In the same way, for each $n \geq 2$,

$$\begin{aligned} A_1 \times \cdots \times A_n &= \{(x_1, \dots, x_n) \mid x_1 \in A_1, \dots, x_n \in A_n\}, \\ A^n &= \{(x_1, \dots, x_n) \mid x_1, \dots, x_n \in A\}. \end{aligned}$$

We call $A_1 \times \cdots \times A_n$ the **Cartesian product** of $A_1, \dots, A_n$.

2.16. Lemma. (1) If $A_1, \dots, A_n$ are all countable, so is their Cartesian product $A_1 \times \cdots \times A_n$.

(2) For every countable set A , each A^n ($n \geq 2$) and the union

$$\bigcup_{n=2}^{\infty} A^n = \{(x_1, \dots, x_n) \mid n \geq 2, x_1, \dots, x_n \in A\}$$

are all countable.

Proof. (1) If some A_i is empty, then the product is empty (by the definition) and hence countable. Otherwise, in the case of two sets A, B , we have some enumeration

$$B = \{b_0, b_1, \dots\}$$

of B , obviously

$$A \times B = \bigcup_{n=0}^{\infty} (A \times \{b_n\}),$$

and each $A \times \{b_n\}$ is equinumerous with A (and hence countable) via the correspondence $(x \mapsto (x, b_n))$.

²In "mathematical English," when we say "for any two objects x, y ," we do not mean that necessarily $x \neq y$, e.g. the assertion that "for any two numbers x, y , $(x + y)^2 = x^2 + 2xy + y^2$ " implies that "for every number x , $(x + x)^2 = x^2 + 2xx + x^2$."

(2) follows by induction for products of n factors and for $\bigcup_{n=2}^{\infty} A^n$ we appeal once more to **2.10**. $\dashv$

2.17. Definition. A real number α is **algebraic** if it is a root of some polynomial

$$P(x) = a_0 + a_1x + \cdots + a_nx^n$$

with integer coefficients $a_0, \dots, a_n \in \mathbb{Z}$ ($n \geq 1, a_n \neq 0$), i.e. if

$$P(\alpha) = 0.$$

Typical examples of algebraic numbers are $\sqrt{2}$, $(1 + \sqrt{2})^2$ (why?) but also the real root of the equation $x^5 + x + 1 = 0$ which exists (why?) but cannot be expressed in terms of radicals, by a classical theorem of Abel.

2.18. Corollary. The set K of algebraic real numbers is countable (Cantor), and hence there exist real numbers which are not algebraic (Liouville).

Proof. The set Π of all polynomials with integer coefficients is countable, because each such polynomial is determined by the sequence of its coefficients, so that Π can be injected into the countable set $\bigcup_{n=2}^{\infty} \mathbb{Z}^n$. For each polynomial $P(x)$, the set of its roots

$$\Lambda(P(x)) = \{\alpha \mid P(\alpha) = 0\}$$

is finite and hence countable. It follows that the set of algebraic numbers K is the union of a sequence of countable sets and hence it is countable. $\dashv$

This first application of the (then) new theory of sets was instrumental in ensuring its quick and favorable acceptance by the mathematicians of the period, particularly since the earlier proof of Liouville (that there exist non-algebraic numbers) was quite intricate. Cantor showed something stronger, that “almost all” real numbers are not algebraic, and he did it with a much simpler proof which used essentially nothing but the completeness of $\mathcal{R}$.

So far we have shown the existence of only two “orders of infinity,” that of N —the countable, infinite sets—and that of $\mathcal{R}$. There are many others.

2.19. Definition. The **powerset** $\mathcal{P}(A)$ of a set A is the set of all its subsets,

$$\mathcal{P}(A) = \{X \mid X \text{ is a set and } X \subseteq A\}.$$

2.20. Exercise. For all sets A, B ,

$$A =_c B \implies \mathcal{P}(A) =_c \mathcal{P}(B).$$

2.21. Theorem. (Cantor) *For every set A ,*

$$A <_c \mathcal{P}(A),$$

i.e. $A \leq_c \mathcal{P}(A)$ but $A \neq_c \mathcal{P}(A)$.

Proof. That $A \leq_c \mathcal{P}(A)$ follows from the fact that the function

$$(x \mapsto \{x\})$$

which associates with each member x of A its **singleton** $\{x\}$ is an injection. (Careful here: the singleton $\{x\}$ is a set with just the one member x and it is not the same object as x , which is probably not a set to start with!)

To complete the proof, we assume (towards a contradiction) that there exists a correspondence

$$\pi : A \twoheadrightarrow \mathcal{P}(A)$$

which witnesses that $A =_c \mathcal{P}(A)$ and we define the set

$$B = \{x \in A \mid x \notin \pi(x)\}.$$

Now B is a subset of A and π is a surjection, so there must exist some $b \in A$ such that $B = \pi(b)$, and (as for each $x \in A$) *either* $b \in B$ *or* $b \notin B$. If $b \in B$, then $b \in \pi(b)$ since $B = \pi(b)$, so that b does not satisfy the condition which defines B , and hence $b \notin B$, contrary to hypothesis. If $b \notin B$, then $b \notin \pi(b)$, so that b now satisfies the defining condition for B and hence $b \in B$, which again contradicts the hypothesis. Thus we reach a contradiction from the assumption that the bijection π exists and the proof is complete. $\dashv$

REMARK. We have actually shown the somewhat stronger proposition, that for each A *there is no surjection* $\pi : A \twoheadrightarrow \mathcal{P}(A)$. A careful examination will reveal that this proof is a fairly straightforward generalization of the second diagonal method of proof by which Cantor showed that the set Δ of infinite binary sequences is uncountable.

So there are many orders of infinity, and specifically those of the sets

$$N <_c \mathcal{P}(N) <_c \mathcal{P}(\mathcal{P}(N)) <_c \cdots.$$

If we name these sets by the recursion

$$\begin{aligned} T_0 &= N, \\ T_{n+1} &= \mathcal{P}(T_n), \end{aligned} \tag{2.1}$$

then their union $T_\infty = \bigcup_{n=0}^\infty T_n$ has a larger cardinality than each T_n , Problem *x2.5. The classification and study of these orders of infinity is one of the central problems of set theory.

The next obvious problem is the comparison for size of the two simplest uncountable sets, the real numbers $\mathcal{R}$ and the powerset $\mathcal{P}(N)$ of the natural numbers.

2.22. Lemma. $\mathcal{P}(N) \leq_c \mathcal{R}$.

Proof. It is enough to prove that $\mathcal{P}(N) \leq_c \Delta$, since we have already shown that $\Delta \leq_c \mathcal{R}$. This, however, is obvious by the map $(X \mapsto c_X)$ which associates with each $X \subseteq N$ the binary sequence

$$c_X(i) = \begin{cases} 1, & \text{if } i \in X, \\ 0, & \text{if } i \notin X, \end{cases}$$

and which is an injection because if some i belongs to one of the sets X, Y and not the other, then $c_X(i) \neq c_Y(i)$. $\dashv$

2.23. Lemma. $\mathcal{R} \leq_c \mathcal{P}(N)$.

Proof. It is enough to show that $\mathcal{R} \leq_c \mathcal{P}(Q)$, since the set of rationals Q is equinumerous with N and hence $\mathcal{P}(N) =_c \mathcal{P}(Q)$. This follows from the fact that the function

$$x \mapsto \pi(x) = \{q \in Q \mid q < x\} \subseteq Q$$

is an injection, because if $x < y$ are distinct real numbers, then there exists some rational q between them, $x < q < y$ and $q \in \pi(y) \setminus \pi(x)$. $\dashv$

From these simple Lemmas it follows that the equinumerosity $\mathcal{R} =_c \mathcal{P}(N)$ will be a direct Corollary of the following basic theorem.

2.24. Schröder-Bernstein Theorem. *For any two sets A, B ,*

$$A \leq_c B \ \& \ B \leq_c A \implies A =_c B.$$

Proof.³ We assume that there exist injections

$$f : A \hookrightarrow B, \quad g : B \hookrightarrow A,$$

and we define the sets A_n, B_n by the following recursion:

$$\begin{aligned} A_0 &= A, & B_0 &= B, \\ A_{n+1} &= gf[A_n], & B_{n+1} &= fg[B_n], \end{aligned}$$

where $fg[X] = \{f(g(x)) \mid x \in X\}$ and correspondingly for the function gf . By induction on n (easily)

$$\begin{aligned} A_n &\supseteq g[B_n] \supseteq A_{n+1}, \\ B_n &\supseteq f[A_n] \supseteq B_{n+1}, \end{aligned}$$

³A different proof of this theorem is outlined in Problems *x4.26, *x4.27.

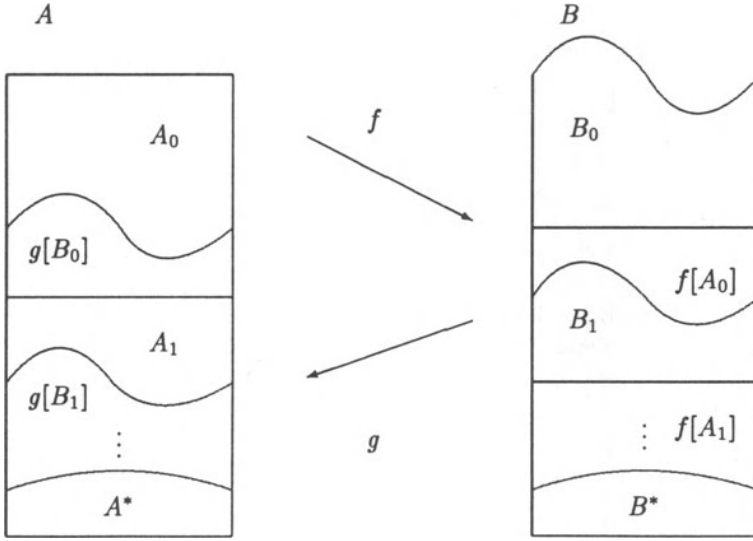


Figure 2.5. Proof of the Schröder-Bernstein Theorem.

so that we have the “chains of inclusions”

$$\begin{aligned} A_0 \supseteq g[B_0] \supseteq A_1 \supseteq g[B_1] \supseteq A_2 \cdots, \\ B_0 \supseteq f[A_0] \supseteq B_1 \supseteq f[A_1] \supseteq B_2 \cdots. \end{aligned}$$

We also define the intersections

$$A^* = \bigcap_{n=0}^{\infty} A_n, \quad B^* = \bigcap_{n=0}^{\infty} B_n,$$

so that

$$B^* = \bigcap_{n=0}^{\infty} B_n \supseteq \bigcap_{n=0}^{\infty} f[A_n] \supseteq \bigcap_{n=0}^{\infty} B_{n+1} = B^*$$

and since f is an injection, by Problem **x1.6**

$$f[A^*] = f[\bigcap_{n=0}^{\infty} A_n] = \bigcap_{n=0}^{\infty} f[A_n] = B^*.$$

Thus f is a bijection of A^* with B^* . On the other hand,

$$A = A^* \cup (A_0 \setminus g[B_0]) \cup (g[B_0] \setminus A_1) \cup (A_1 \setminus g[B_1]) \cup (g[B_1] \setminus A_2) \cdots$$

$$B = B^* \cup (B_0 \setminus f[A_0]) \cup (f[A_0] \setminus B_1) \cup (B_1 \setminus f[A_1]) \cup (f[A_1] \setminus B_2) \cdots$$

and these sequences are separated, i.e. no set in them has any common element with any other. To finish the proof it is enough to check that for every n ,

$$\begin{aligned} f[A_n \setminus g[B_n]] &= f[A_n] \setminus B_{n+1}, \\ g[B_n \setminus f[A_n]] &= g[B_n] \setminus A_{n+1}, \end{aligned}$$

from which the first (for example) is true because f is an injection and

$$f[A_n \setminus g[B_n]] = f[A_n] \setminus fg[B_n] = f[A_n] \setminus B_{n+1}.$$

Finally we have the bijection $\pi : A \rightarrow B$,

$$\pi(x) = \begin{cases} f(x), & \text{if } x \in A^* \text{ or } (\exists n)[x \in A_n \setminus g[B_n]], \\ g^{-1}(x), & \text{if } x \notin A^* \text{ and } (\exists n)[x \in g[B_n] \setminus A_{n+1}], \end{cases}$$

which verifies that $A =_c B$ and finishes the proof. $\dashv$

Using the Schröder-Bernstein Theorem we can establish easily several equinumerosities which are quite difficult to prove directly.

Problems

x2.1. For every $\alpha < \beta$ where α, β are reals, ∞ or $-\infty$, construct bijections which prove the equinumerosities

$$(\alpha, \beta) =_c (0, 1) =_c \mathcal{R}.$$

x2.2. For every $\alpha < \beta$, construct bijections which prove the equinumerosities

$$[\alpha, \beta] =_c [\alpha, \beta] =_c \mathcal{R}.$$

x2.3. $\mathcal{P}(N) =_c \mathcal{R} =_c \mathcal{R}^n$, for every $n \geq 2$.

2.25. Definition. For any two sets A, B ,

$$\begin{aligned} (A \rightarrow B) &=_{\text{df}} \{f \mid f : A \rightarrow B\} \\ &= \text{the set of all functions from } A \text{ to } B. \end{aligned}$$

x2.4. For any three sets A, B, C ,

$$((A \times B) \rightarrow C) =_c (A \rightarrow (B \rightarrow C)).$$

***x2.5.** Using the definition (2.1), for every m ,

$$T_m <_c T_\infty = \bigcup_{n=0}^{\infty} T_n.$$

You need to know something about continuous functions to do the last two problems.

***x2.6.** The set $C[0, 1]$ of all continuous, real functions on the closed interval $[0, 1]$ is equinumerous with $\mathcal{R}$.

***x2.7.** The set of all monotone real functions on the closed interval $[0, 1]$ is equinumerous with $\mathcal{R}$.

Chapter 3

PARADOXES AND AXIOMS

In the preceding chapter we gave a brief exposition of the first, basic results of set theory, as it was created by Cantor and the pioneers who followed him in the last twenty five years of the 19th century. By the beginning of our own century, the theory had matured and justified itself with diverse and significant applications, particularly in mathematical analysis. Perhaps its greatest success was the creation of an exceptionally beautiful and useful *transfinite arithmetic*, which introduces and studies the operations of addition, multiplication and exponentiation on infinite numbers. By 1900, there were still two fundamental problems about equinumerosity which remained unsolved. These have played a decisive role in the subsequent development of set theory and we will consider them carefully in the following chapters. Here we just state them, in the form of hypotheses.

3.1. Hypothesis of Cardinal Comparability. *For any two sets A , B , either $A \leq_c B$ or $B \leq_c A$.¹*

3.2. Continuum Hypothesis. *There is no set of real numbers X with cardinality intermediate between those of N and $\mathcal{R}$, i.e.*

$$(\text{CH}) \quad (\forall X \subseteq \mathcal{R})[X \leq_c N \vee X =_c \mathcal{R}].$$

Since $\mathcal{R} =_c \mathcal{P}(N)$, **CH** is a special case of the **Generalized Continuum Hypothesis**, the statement that for every infinite set A ,

$$(\text{GCH}) \quad (\forall X \subseteq A)[X \leq_c A \vee X =_c \mathcal{P}(A)].$$

One immediate consequence of these two hypotheses is that the integers N and the reals $\mathcal{R}$ represent the smallest two “orders of infinity”; every infinite set is either countable, equinumerous with $\mathcal{R}$ or strictly greater than $\mathcal{R}$ in cardinality.

¹Cantor announced the “theorem of comparability of cardinals” in 1895 and in 1899 he outlined a proposed proof of it in a letter to Dedekind, which was not, however, published until 1932. There were problems with that argument and it is probably closer to the truth to say that until 1900 (at least) the question of comparability of cardinals was still open.

In this beginning “naive” phase, set theory was developed on the basis of Cantor’s definition of sets quoted in Chapter 1, much as we proved its basic results in Chapter 2. If we analyze carefully the proofs of those results we will see that they are all based on the following simple principle.

3.3. General Comprehension Principle. *For each n -ary definite condition P , there is a set*

$$A = \{\vec{x} \mid P(\vec{x})\}$$

whose members are precisely all the n -tuples of objects which satisfy $P(\vec{x})$, so that for all $\vec{x}$,

$$\vec{x} \in A \iff P(\vec{x}). \quad (3.1)$$

The extensionality principle implies that at most one set A can satisfy (3.1), and we call this A the **extension** of the condition P .

3.4. Definite conditions and operations. It is necessary to restrict the comprehension principle to definite conditions to avoid questions of vagueness which have nothing to do with science. We do not want to admit the “set”

$$A =_{\text{df}} \{x \mid x \text{ is an honest politician}\},$$

because membership of some specific public figure in it may be a hotly debated topic. *An n -ary condition P is definite if for each n -tuple of objects $\vec{x} = (x_1, \dots, x_n)$, it is determined unambiguously whether $P(\vec{x})$ is true or false.* For example, the binary conditions

$$\begin{aligned} P(x, y) &\iff_{\text{df}} x \text{ is a parent of } y, \\ S(s, t) &\iff_{\text{df}} s \text{ and } t \text{ are siblings} \\ &\iff (\exists x)[P(x, s) \ \& \ P(x, t)] \end{aligned}$$

are both definite, assuming (for the example) that the laws of biology determine parenthood unambiguously. The General Comprehension Principle applies to them and we can form the sets of pairs

$$\begin{aligned} A &=_{\text{df}} \{(x, y) \mid x \text{ is a parent of } y\}, \\ B &=_{\text{df}} \{(s, t) \mid s \text{ and } t \text{ are siblings}\}. \end{aligned}$$

We do not demand of a definite condition that its truth value be *effectively determined*. For example, it is a famous open problem of number theory whether there exist infinitely many pairs of successive, odd primes, and the truth or falsity of the condition

$$G(n) \iff_{\text{df}} n \in N \ \& \ (\exists m > n)[m, m + 2 \text{ are both prime numbers}]$$

is not known for sufficiently large n . Still the condition G is unambiguous and we can use it to form the set of numbers

$$C =_{\text{df}} \{n \in N \mid (\exists m > n)[m, m + 2 \text{ are both prime numbers}]\}.$$

The *twin prime conjecture* asserts that $G = N$, but if it is false, then C is some large, initial segment of the natural numbers.

In the same way, an n -ary operation F is **definite** if it assigns to each n -tuple of objects $\vec{x}$ a unique, unambiguously determined object $w = F(\vec{x})$. For example, assuming again that biology will not betray us, the operation

$$F(x) =_{\text{df}} \begin{cases} \text{the father of } x, & \text{if } x \text{ is a human,} \\ x, & \text{otherwise,} \end{cases}$$

is definite. The silly consideration of cases here was put in to ensure that F determines a value for each argument x . In practice, we would define this operation by the simpler

$$F(x) =_{\text{df}} \text{the father of } x,$$

leaving it to the reader to supply some conventional, irrelevant value $F(x)$ for non-human x 's. Again, definite operations need not be *effectively computable*, in fact the determination of the value $F(x)$ is sometimes the subject of judicial conflict in this specific case.

In addition to the General Comprehension Principle, we also assumed in the preceding chapter the existence of some specific sets, including the sets N and $\mathcal{R}$ of natural and real numbers, as well as the definiteness of some basic conditions from classical mathematics, e.g. the condition of "being a function,"

$$\text{Function}(f, A, B) \iff f \text{ is a function from } A \text{ to } B.$$

This poses no problem as mathematicians have always made these assumptions, explicitly or implicitly.

The General Comprehension Principle has such strong intuitive appeal that the next theorem is called a "paradox."

3.5. Russell's paradox. *The General Comprehension Principle is not valid.*

Proof. Notice first that if the General Comprehension Principle holds, then the set of all sets

$$V =_{\text{df}} \{x \mid x \text{ is a set}\}$$

is a set, and it has the somewhat peculiar property that it belongs to itself, $V \in V$. The common sets of everyday mathematics—sets of numbers, functions, etc.—surely do not contain themselves, so it is natural to consider them as members of a smaller, more natural universe of sets, by applying the General Comprehension Principle again,

$$R = \{x \mid x \text{ is a set and } x \notin x\}.$$

From the definition of R , however,

$$R \in R \iff R \notin R,$$

which is absurd.

When it is more than just a mistake, a “paradox” is simply a fact which runs counter to our intuitions, and set theorists already knew several such “paradoxes” before Russell announced this one in 1902, in a historic letter to the leading German philosopher and founder of mathematical logic Gottlob Frege. These other paradoxes, however, were technical and affected only some of the most advanced parts of Cantor’s theory. One could imagine that higher set theory had a systematic error built in, something like allowing a careless “division by 0” which would soon be discovered and disallowed, and then everything would be fixed. After all, contradictions and paradoxes had plagued the “infinitesimal calculus” of Newton and Leibnitz and they all went away after the rigorous foundation of the theory which was just being completed in the 1890s, without affecting the vital parts of the subject. Russell’s paradox, however, was something else again: simple and brief, it affected directly the fundamental notion of set and the “obvious” principle of comprehension on which set theory had been built. It is not an exaggeration to say that Russell’s paradox brought a foundational *crisis of doubt*, first to set theory and through it, later, to all of mathematics, which took over thirty years to overcome.

Some, like the French geometer Poincare and the Dutch topologist and philosopher Brouwer, proposed radical solutions which essentially dismissed set theory (and much of classical mathematics along with it) as “pseudotheries,” without objective content. From those who were reluctant to leave “Cantor’s paradise,” Russell first attempted to “rescue” set theory with his famous *theory of types*, which, however, is awkward to apply and was not accepted by a majority of mathematicians.² At approximately the same time, Zermelo proposed an alternative solution, which in time and with the contributions of many evolved into the contemporary theory of sets.

In his first publication on the subject in 1908, Zermelo took a pragmatic view of the problem. No doubt the General Comprehension Principle was not generally valid, Russell’s paradox had made that clear. On the other hand, the specific applications of this principle in the proofs of basic facts about sets (like those in Chapter 2) are few, simple and seemingly non-contradictory.

²The theory of types had a strong influence in the development of analytic philosophy and logic in our century and some of its basic ideas eventually found their place in set theory also.

Under such circumstances there is at this point nothing left for us to do but to proceed in the opposite direction [from that of the General Comprehension Principle] and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all that is valuable in this theory.

In other words, Zermelo proposed to replace the direct *intuitions* of Cantor about sets which led us to the faulty general comprehension principle with some *axioms*, hypotheses about sets which we accept with little a priori justification, simply because they are necessary for the proofs of the fundamental results of the existing theory and seemingly free of contradiction.

Such were the philosophically dubious beginnings of **axiomatic set theory**, surely one of the most significant achievements of 20th century science. From its inception, however, the new theory had a substantial advantage in the genius of Zermelo, who selected an extraordinarily natural and pliable axiomatic system. None of Zermelo's axioms has yet been discarded or seriously revised and (until very recently) only one basic new axiom was added to his seven in the decade 1920-1930. In addition, despite the opportunistic tone of the cited quotation, each of Zermelo's axioms expresses a property of sets which is intuitively obvious and was already well understood from its uses in classical mathematics. With the experience gained from working out the consequences of these axioms over the years, a new intuitive notion of "grounded set" has been created which does not lead to contradictions and for which the axioms of set theory are clearly true. We will reconsider the problem of foundation of set theory after we gain experience by the study of its basic mathematical results.

The basic model for the axiomatization of set theory was Euclidean geometry, which for 2000 had been considered the "perfect" example of a rigorous, mathematical theory. If nothing else, the axiomatic method clears the waters and makes it possible to separate what might be confusing and self-contradictory in our intuitions about the objects we are studying, from simple errors in logic we might be making in our proofs. As we proceed in our study of axiomatic set theory, it will be useful to remind ourselves occasionally of the example of Euclidean geometry.

3.6. The axiomatic setup. We assume at the outset that there is a **domain** or **universe** $\mathcal{W}$ of **objects**, some of which are **sets**, and certain **definite conditions and operations** on $\mathcal{W}$, among them the basic conditions of *identity*, *sethood* and *membership*:

$$x = y \iff x \text{ is the same object as } y,$$

$$\begin{aligned} \text{Set}(x) &\iff x \text{ is a set,} \\ x \in y &\iff \text{Set}(y) \text{ and } x \text{ is a member of } y. \end{aligned}$$

We call the objects which are not sets **atoms**, but we do not require that any atoms exist, i.e. it may be the case that all the objects are sets.

This is the way every axiomatic theory begins. In Euclidean geometry for example, we start with the assumption that there are *points*, *lines* and several other geometrical objects and that some basic, definite conditions and operations are specified on them, e.g. it makes sense to ask if a “point P lies on the line L ,” or “to construct a line joining two given points.” We then proceed to formulate the classical axioms of Euclid about these objects and to derive theorems from them. Actually Euclidean geometry is quite complex: there are several types of basic objects and a long list of intricate axioms about them. By contrast, Zermelo’s set theory is quite austere: we just have sets and atoms and only seven fairly simple axioms relating them. In the remainder of this chapter we will introduce six of these axioms with a few comments and examples. It is a bit easier to put off stating his last, seventh axiom until we first gain some understanding of the consequences of the first six in the next few chapters.

3.7. (I) Axiom of Extensionality. *For any two sets A , B ,*

$$A = B \iff (\forall x)[x \in A \iff x \in B].$$

3.8. (II) Emptyset and Pairset Axioms. (a) *There is a special object $\emptyset$ which we will call a set, but which has no members.* (b) *For any two objects x , y , there is a set A whose only members are x and y , so that it satisfies the equivalence*

$$t \in A \iff t = x \vee t = y. \quad (3.2)$$

The Axiom of Extensionality implies that only one empty set exists, and that for any two objects x , y , only one set A can satisfy (3.2). We denote this **doubleton** of x and y by

$$\{x, y\} =_{\text{df}} \text{the unique set } A \text{ with sole members } x, y.$$

If $x = y$, then $\{x, x\} = \{x\}$ is the **singleton** of the object x .

Using this axiom we can construct many simple sets, e.g.

$$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}, \{\{\emptyset\}, \{\{\emptyset\}\}\}, \dots,$$

but each of them has at most two members!

3.9. Exercise. *Prove that $\emptyset \neq \{\emptyset\}$.*

3.10. (III) Separation Axiom or Axiom of Subsets. *For each set A and each unary, definite condition P , there exists a set B which satisfies the equivalence*

$$x \in B \iff x \in A \ \& \ P(x). \quad (3.3)$$

From the Extensionality Axiom again, it follows that only one B can satisfy (3.3) and we will denote it by

$$B = \{x \in A \mid P(x)\}.$$

A characteristic contribution of Zermelo, this axiom is obviously a restriction of the General Comprehension Principle which implies many of its trouble-free consequences. For example, we can use it to define the operations of intersection and difference on sets,

$$\begin{aligned} A \cap B &=_{\text{df}} \{x \in A \mid x \in B\}, \\ A \setminus B &=_{\text{df}} \{x \in A \mid x \notin B\}. \end{aligned}$$

The proof of Russell's paradox yields a theorem:

3.11. Theorem. *For each set A , the set*

$$\mathbf{r}(A) =_{\text{df}} \{x \in A \mid x \notin x\} \quad (3.4)$$

is not a member of A . It follows that the collection of all sets is not a set, i.e. there is no set V which satisfies the equivalence

$$x \in V \iff \text{Set}(x).$$

Proof. Notice first that $\mathbf{r}(A)$ is a set by the Separation Axiom. Assuming that $\mathbf{r}(A) \in A$, we have (as before) the equivalence

$$\mathbf{r}(A) \in \mathbf{r}(A) \iff \mathbf{r}(A) \notin \mathbf{r}(A),$$

which is absurd. ⊥

3.12. (IV) Powerset Axiom. *For each object A , there exists a set B whose members are the subsets of A , i.e.*

$$X \in B \iff \text{Set}(X) \ \& \ X \subseteq A. \quad (3.5)$$

Here $X \subseteq A$ is an abbreviation of $(\forall t)[t \in X \implies t \in A]$. The Axiom of Extensionality implies that for each A , only one set can satisfy (3.5), we call it the **powerset** of A and we denote it by

$$\mathcal{P}(A) =_{\text{df}} \{X \mid \text{Set}(X) \ \& \ X \subseteq A\}.$$

3.13. Exercise. If A is an atom or $A = \emptyset$, then $\mathcal{P}(A) = \{\emptyset\}$.

3.14. Exercise. For each set A , there exists a set B whose members are exactly all singletons of members of A , i.e.

$$x \in B \iff (\exists t \in A)[x = \{t\}].$$

3.15. (V) Unionset Axiom. For every object $\mathcal{E}$, there exists a set B whose members are the members of the members of $\mathcal{E}$, i.e. it satisfies the equivalence

$$t \in B \iff (\exists X \in \mathcal{E})[t \in X]. \quad (3.6)$$

The Axiom of Extensionality implies again that for each $\mathcal{E}$, only one set can satisfy (3.6), we call it the **unionset** of $\mathcal{E}$ and we denote it by

$$\bigcup \mathcal{E} =_{\text{df}} \{t \mid (\exists X \in \mathcal{E})[t \in X]\}.$$

The unionset operation is obviously most useful when $\mathcal{E}$ is a **family of sets**, i.e. when $\mathcal{E}$ and each $X \in \mathcal{E}$ are sets. This is the case for the simplest application, which (finally) gives us the binary, union operation on sets: we set

$$A \cup B = \bigcup \{A, B\}$$

using axioms (II) and (V), and we compute

$$\begin{aligned} t \in A \cup B &\iff (\exists X \in \{A, B\})[t \in X] \\ &\iff t \in A \vee t \in B. \end{aligned}$$

It is convenient, however, to have $\bigcup \mathcal{E}$ defined for arbitrary objects $\mathcal{E}$.

3.16. Exercise. If $\mathcal{E}$ is an atom, then $\bigcup \mathcal{E} = \emptyset$ and $\bigcup \emptyset = \bigcup \{\emptyset\} = \emptyset$.

3.17. (VI) Axiom of Infinity. There exists a set I which contains the empty set $\emptyset$ and the singleton of each of its members, i.e.

$$\emptyset \in I \ \& \ (\forall x)[x \in I \implies \{x\} \in I].$$

We have not given yet a rigorous definition of “infinite,” but it is quite obvious that any I with the properties in the axiom must be infinite, since (VI) implies

$$\emptyset \in I, \ \{\emptyset\} \in I, \ \{\{\emptyset\}\} \in I, \dots$$

and the objects $\emptyset, \{\emptyset\}, \dots$ are all distinct sets by the Extensionality Axiom. The intuitive understanding of the axiom is that it demands precisely the existence of the set

$$I = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\},$$

but it is simpler (and sufficient) to assume of I only the stated properties, which imply that it contains all these complex singletons.

It was a commonplace belief among philosophers and mathematicians of the 19th century that the existence of infinite sets could be proved, and in particular the set of natural numbers could be “constructed” out of thin air, “by logic alone.” All the proposed “proofs” involved the faulty General Comprehension Principle in some form or another. We know better now: logic can codify the valid forms of reasoning but it cannot prove the existence of anything, let alone infinite sets. By taking account of this fact cleanly and explicitly in the formulation of his axioms, Zermelo made a substantial contribution to the process of purging logic of ontological concerns, a necessary step in the rigorous development of logic as a science in its own right in our century.

3.18. Axioms for definite conditions and operations. Zermelo understood *definite conditions* intuitively, he described them much as we did in 3.4 and he applied the Separation Axiom using various quite complex conditions without any special argument that they are, indeed, “definite.” We will do the same, because the business of proving definiteness is boring and not particularly illuminating. For the sake of completeness, however, we list here the only properties of definiteness that we will actually use.

1. The following basic conditions are definite:

$$\begin{aligned} x = y &\iff_{\text{df}} x \text{ and } y \text{ are the same object,} \\ \text{Set}(x) &\iff_{\text{df}} x \text{ is a set,} \\ x \in y &\iff_{\text{df}} \text{Set}(y) \text{ and } x \text{ is a member of } y, \end{aligned}$$

2. For each object c and each n , the constant n -ary operation

$$F(x_1, \dots, x_n) = c$$

is definite.

3. Each *projection operation*

$$F(x_1, \dots, x_n) = x_i \quad (1 \leq i \leq n)$$

is definite.

4. If P is a definite condition of $n + 1$ arguments and for each tuple $\vec{x} = x_1, \dots, x_n$ there exists exactly one w such that $P(\vec{x}, w)$, then the operation

$$F(\vec{x}) = \text{the unique } w \text{ such that } P(\vec{x}, w)$$

is definite.

5. If Q is an m -ary definite condition, each F_i is an n -ary definite operation for $i = 1, \dots, m$ and

$$P(\vec{x}) \iff_{\text{df}} Q(F_1(\vec{x}), \dots, F_m(\vec{x})),$$

then the condition P is also definite.

6. If Q , R and S are definite conditions of the appropriate number of arguments, then so are the following conditions which are obtained from them by applying the elementary operations of logic:

$$\begin{array}{llll} P_1(\vec{x}) & \iff_{\text{df}} & \neg P(\vec{x}) & \iff P(\vec{x}) \text{ is false,} \\ P_2(\vec{x}) & \iff_{\text{df}} & Q(\vec{x}) \ \& \ R(\vec{x}) & \iff \text{both } Q(\vec{x}) \text{ and } R(\vec{x}) \text{ are true,} \\ P_3(\vec{x}) & \iff_{\text{df}} & Q(\vec{x}) \vee R(\vec{x}) & \iff \text{either } Q(\vec{x}) \text{ or } R(\vec{x}) \text{ is true,} \\ P_4(\vec{x}) & \iff_{\text{df}} & (\exists y)S(\vec{x}, y) & \iff \text{for some } y, S(\vec{x}, y) \text{ is true,} \\ P_5(\vec{x}) & \iff_{\text{df}} & (\forall y)S(\vec{x}, y) & \iff \text{for every } y, S(\vec{x}, y) \text{ is true.} \end{array}$$

All the conditions and operations we will use can be proved definite by appealing to these basic properties. Aside from one problem at the end of this chapter, however, for the logically minded, we will omit these technical proofs of definiteness and it is best for the reader to forget about them too: they detract from the business at hand, which is the study of sets, not definite conditions and operations.

3.19. Classes. Having gone to all the trouble to discredit the General Principle of Comprehension, we will now profess that *for every unary, definite condition P there exists a class*

$$A = \{x \mid P(x)\}, \quad (3.7)$$

such that for every object x ,

$$x \in A \iff P(x). \quad (3.8)$$

To give meaning to this principle and prove it, we need a simple notational convention, and the important notion of a “class.” Every set will be a class, but because of the Russell Paradox **3.5**, there must be more classes than sets, else (3.7) and (3.8) lead immediately to a contradiction.

First let us agree that for every unary, definite condition P we will write synonymously

$$x \in P \iff P(x).$$

For example, if *Set* is the basic condition of sethood, we write interchangeably

$$x \in \text{Set} \iff \text{Set}(x) \iff x \text{ is a set.}$$

This is just a useful notation.

A unary definite condition P is **coextensive** with a set A if the objects which satisfy it are precisely the members of A ,

$$P =_e A \iff_{\text{df}} (\forall x)[P(x) \iff x \in A]. \quad (3.9)$$

For example, if

$$P(x) \iff x \neq x,$$

then $P =_e \emptyset$. By the Russell Paradox 3.5, not every P is coextensive with a set. On the other hand, a unary, definite condition P is coextensive with at most one set; because if $P =_e A$ and also $P =_e B$, then for every x ,

$$x \in A \iff P(x) \iff x \in B,$$

and $A = B$ by the Axiom of Extensionality.

A **class** A is either a set or a unary definite condition which is not coextensive with a set. With each unary P , we associate the class

$$\{x \mid P(x)\} =_{\text{df}} \begin{cases} \text{the unique set } A \text{ such that } P =_e A, \\ \text{if } P =_e A \text{ for some set } A, \\ P, \text{ otherwise.} \end{cases}$$

Now if $A =_{\text{df}} \{x \mid P(x)\}$, then either P is coextensive with a set, in which case $P =_e A$ and by the definition $x \in A \iff P(x)$; or P is not coextensive with any set, in which case A is a definite condition and

$$\begin{aligned} x \in A &\iff A(x) \text{ by the notational convention,} \\ &\iff P(x) \text{ because } A = P. \end{aligned}$$

This is exactly the *General Comprehension Principle for Classes* enunciated above.

3.20. Exercise. For every set A ,

$$\{x \mid x \in A\} = A,$$

and, in particular, every set is a class. Show also that

$$\{X \mid \text{Set}(X) \ \& \ X \subseteq A\} = \mathcal{P}(A).$$

3.21. Exercise. The class of all singletons $\{X \mid (\exists y)[X = \{y\}]\}$ is not a set.

3.22. Exercise. For every class A ,

$$\begin{aligned} A \text{ is a set} &\iff \text{for some class } B, A \in B \\ &\iff \text{for some set } X, A \subseteq X, \end{aligned}$$

where inclusion among classes is defined as if they were sets,

$$A \subseteq B \iff_{\text{df}} (\forall x)[x \in A \implies x \in B].$$

3.23. The Axioms of Choice and Replacement: a warning. Our axiomatization of set theory will not be complete until we introduce Zermelo's last Axiom of Choice in Chapter 8 and the later Axiom of Replacement in Chapter 11. While there are good reasons for these postponements which we will explain in due course, there are also good reasons for adding the axioms of Choice and Replacement: many basic set theoretic arguments need them, and among these are some of the simplest claims of Chapter 2. Thus, until Chapter 8, we will need to be extra careful and make sure that our constructions indeed can be justified by axioms (I) - (VI) and that we have not sneaked in some "obvious" assertion about sets not yet proved or assumed. In a few places we will formulate and prove something weaker than the whole truth whose proof happens to need one of the missing axioms. Now this is good: it will keep us on our toes and make us understand better the art of reasoning from axioms.

3.24. About atoms. Most recent developments of axiomatic set theory assume at the outset the so-called **Principle of Purity**, that *there are no atoms*, all objects of the basic domain are *sets*. There is a certain appealing simplicity to this conception of a mathematical world in which everything is a set. We have followed Zermelo in allowing atoms (without demanding them), primarily because this makes the theory more naturally applicable to general mathematics and science. In any case, it comes at little cost, we simply have to say "object" in some situations where the atom banners would say "set." It is important to notice, however, that none of the axioms requires the existence of atoms, so none of the consequences we will derive from them depends on the existence of atoms: everything we will prove remains true in the domain of pure sets, provided only that it satisfies the Zermelo axioms, as we stated them.

3.25. Axioms as closure properties of the universe $\mathcal{W}$. Whatever the domain $\mathcal{W}$ of our axiomatic set theory may be, it is clear that it does not contain all "objects of our intuition or thought" in Cantor's expression; $\mathcal{W}$ is not a set, and it is certainly a perfectly legitimate mathematical object of our intuition about which we intend to have many thoughts. Granting that $\mathcal{W}$ is not all there is, we can fruitfully conceive of the axioms as imposing *closure conditions* on it. We have assumed (so far) that $\mathcal{W}$ contains $\emptyset$, that it is closed under the operations of pairing $\{x, y\}$, (II), powerset $\mathcal{P}(X)$, (IV) and unionset $\bigcup \mathcal{A}$, (V), that it includes every definite subcollection of every set, (III), and that it contains some set I with the stipulated property of the Axiom of Infinity, (VI). It is also possible to understand the Axiom of Extensionality (I) as a closure property of $\mathcal{W}$: in its non-trivial direction,

it says that for any two sets A, B ,

$$A \neq B \implies (\exists t)[t \in (A \setminus B) \cup (B \setminus A)], \quad (3.10)$$

i.e. every inequality $A \neq B$ between two sets is witnessed by some legitimate object $t \in \mathcal{W}$ which belongs to one and not the other.

This understanding of the meaning of the axioms is compatible with two different conceptions of the universe $\mathcal{W}$. One is that it is huge, amorphous, difficult to understand and impossible to define; but every object in it is concrete, definite, whole, and this is enough to justify the closure properties of $\mathcal{W}$ embodied by the axioms. Let us call this *the large view*. The *small view* is that $\mathcal{W}$ consists precisely of those objects whose existence is “guaranteed” by the axioms, those which can be “constructed” by applying the axioms repeatedly: the axioms are satisfied because we deliberately put in $\mathcal{W}$ all the objects required by the closure properties they express. Neither conception is precise, to be sure, but they are different. On the small view, for example, there are no atoms, since none of the axioms demands their existence, while the large view clearly allows lots of them.

Both of these views can be defended and they have played significant roles in the philosophy of set theory, and even in its mathematical practice, by suggesting the kind of questions one should ask, for example. We will come back to discuss the issue in Chapter 11 and Appendix B, when we will be in a position to be less flippant about it. In the meantime, we will often speak of the axioms as closure conditions on $\mathcal{W}$, a useful heuristic device which is compatible with every philosophical approach to the subject.

Problems

x3.1. For each non-empty set $\mathcal{E}$ and each $X \in \mathcal{E}$, we define the *intersection of $\mathcal{E}$ via X* by

$$\bigcap_X \mathcal{E} =_{\text{df}} \{x \in X \mid (\forall U \in \mathcal{E})[x \in U]\}.$$

Show that for any two members X, Y of $\mathcal{E}$,

$$\bigcap_X \mathcal{E} = \bigcap_Y \mathcal{E},$$

i.e. the intersection $\bigcap_X \mathcal{E}$ is independent of the specific X we used in its definition, and hence we can use for it the notation $\bigcap \mathcal{E}$ which does not exhibit X . Show also that $A \cap B = \bigcap \{A, B\}$.

3.26. Definition. Two sets A, B are **disjoint** if $A \cap B = \emptyset$. A set W is a **connection** of the two disjoint sets A and B (according to Zermelo) if the following three conditions hold:

1. $Z \in W \implies (\exists x \in A, y \in B)[Z = \{x, y\}]$.
2. For each $x \in A$, there is exactly one $y \in B$ such that $\{x, y\} \in W$.
3. For each $y \in B$, there is exactly one $x \in A$ such that $\{x, y\} \in W$.

x3.2. For any two disjoint sets A, B , the set $\Sigma(A, B)$ of all connections of A with B exists—i.e. there exists a set $\Sigma(A, B)$ such that

$$W \in \Sigma(A, B) \iff W \text{ is a connection of } A \text{ with } B.$$

3.27. Definition. Two sets A, B are **equivalent according to Zermelo** if there exists a third set C disjoint from both of them and connections of A with C and of B with C , in symbols

$$A \sim_Z B \iff (\exists C, W, W')[A \cap C = \emptyset \ \& \ B \cap C = \emptyset \\ \& \ W \in \Sigma(A, C) \ \& \ W' \in \Sigma(B, C)].$$

***x3.3.** The condition of equivalence according to Zermelo has the following properties, for any three sets A, B, C :

$$\begin{aligned} A &\sim_Z A, \\ A \sim_Z B &\implies B \sim_Z A, \\ A \sim_Z B \ \& \ B \sim_Z C &\implies A \sim_Z C. \end{aligned}$$

x3.4. Prove rigorously that the following conditions and operations are definite, using only (I) - (VI) and the axioms in **3.18**. (Here c is some arbitrary object.)

$$\begin{aligned} P_1(x) &\iff_{\text{df}} x \in c, \\ P_1(x, y, z) &\iff_{\text{df}} z \in x, \\ P_1(X, Y) &\iff_{\text{df}} X \subseteq Y, \\ F(x) &=_{\text{df}} \{x\}, \\ F(X, Y) &=_{\text{df}} X \cup Y. \end{aligned}$$

Chapter 4

ARE SETS ALL THERE IS?

Our next goal is to determine whether the basic results of naive set theory in Chapter 2 can be proved on the basis of the axioms of Zermelo. Right at the start we hit a snag: to define the crucial notion of *equinumerosity* we need functions; to define *countable sets* we need the specific set N of natural numbers; the fundamental theorem 2.21 of Cantor is about the set $\mathcal{R}$ of real numbers, etc. Put another way, the results of Chapter 2 are not only about sets, but about points, numbers, functions, Cartesian products and many other mathematical objects which are plainly not sets. Where will we find these objects in the axioms of Zermelo which speak only about sets?

An obvious solution is to assume that these non-sets are among the *atoms* which are allowed by Zermelo's theory and to add axioms which express our basic intuitions about points, numbers, functions, etc. This is possible but awkward and there is a much better solution.

A typical example of the method we will adopt is the "identification" of the (directed) geometric line Π with the set $\mathcal{R}$ of real numbers, via the correspondence which "identifies" each point $P \in \Pi$ with its coordinate $x(P)$ with respect to a fixed choice of an origin O . What is the precise meaning of this "identification"? *Certainly not that points are real numbers*. Men have always had direct geometric intuitions about points which have nothing to do with their coordinates and which existed before Descartes discovered analytic geometry. Every Athenian of the classical period understood the meaning of the sentence

Phaliron is between Piraeus and Sounion along the Saronic coast¹

even though he was (by necessity) ignorant of analytic geometry. In fact, many educated ancient Athenians had an excellent understanding of the Pythagorean Theorem, without knowing how to coordinatize the plane. What we mean by the "identification" of Π with $\mathcal{R}$ is that the correspon-

¹These are seaside suburbs of Athens.

dence $P \mapsto x(P)$ gives a **faithful representation** of Π in $\mathcal{R}$ which allows us to give arithmetic definitions for all the useful geometric notions and to study the mathematical properties of Π **as if points were real numbers**. For example, the quoted sentence above is expressed by the inequalities

$$x(\text{Piraeus}) < x(\text{Phaliron}) < x(\text{Sounion}),$$

assuming that the coordinates increase in the easterly direction. In the same way, we will discover within the universe of sets *faithful representations* of all the mathematical objects we need, and we will study set theory on the basis of the lean axiomatic system of Zermelo **as if all mathematical objects were sets**. The delicate problem in specific cases is to formulate precisely the correct definition of a “faithful representation” and to prove that one such exists.

We consider first the basic (**ordered**) **pair** operation. Intuitively, the pair (x, y) of two objects x and y is the “thing” which has a “first member” x and a “second member” y , and it is different from the unordered pair $\{x, y\}$ since (for example) if $x \neq y$, then $(x, y) \neq (y, x)$ while $\{x, y\} = \{y, x\}$. Thus, the first characteristic property of the ordered pair is the following:

$$\mathbf{4.1.} \quad (x, y) = (x', y') \iff x = x' \ \& \ y = y'.$$

There is a second, perhaps less obvious characteristic property of pairs which makes it possible to define Cartesian products:

4.2. For any two sets A, B , the class

$$A \times B =_{\text{df}} \{(x, y) \mid x \in A \ \& \ y \in B\}$$

is a set.

Thus, the problem of representing the notion of “pair” in set theory takes the following precise form: we must define a definite operation (x, y) such that **4.1** and **4.2** follow from the axioms of Zermelo.

4.3. Lemma. *The Kuratowski pair operation*

$$(x, y) =_{\text{df}} \{\{x\}, \{x, y\}\} \tag{4.1}$$

has properties 4.1 and 4.2.

Proof. 4.1. The direction $\Leftarrow$ is obvious. For the non-trivial direction $\Rightarrow$, we distinguish two cases.

If $x = y$, then $\{x, y\} = \{x, x\} = \{x\}$, the set $(x, y) = \{\{x\}, \{x\}\} = \{\{x\}\}$ is a singleton, hence the set (x', y') which is assumed equal to it is also a

singleton, so that $x' = y'$ and $(x', y') = \{\{x'\}\}$; and since this last singleton is equal to $\{\{x\}\}$, we have $x = x'$ and, hence, also $y = x = x' = y'$.

If $x \neq y$, then the members of (x, y) are the singleton $\{x\}$ and the doubleton $\{x, y\}$, and these must correspond with the members $\{x'\}$ and $\{x', y'\}$ of the equal set (x', y') , so that we must have $\{x\} = \{x'\}$, $\{x, y\} = \{x', y'\}$, and then, immediately, $x = x'$ and $y = y'$.

Proof of 4.2. The condition

$$\text{OrdPair}_{A,B}(z) \iff_{\text{df}} (\exists x \in A)(\exists y \in B)[z = (x, y)]$$

is evidently definite for each fixed A, B , and hence to verify 4.2, it is enough to find for each A, B some set C such that

$$x \in A \ \& \ y \in B \implies (x, y) \in C;$$

using this C , we can then construct the Cartesian product as a set by the Separation Axiom,

$$A \times B =_{\text{df}} \{z \in C \mid \text{OrdPair}_{A,B}(z)\}.$$

We compute:

$$\begin{aligned} x \in A, y \in B &\implies \{x\}, \{x, y\} \subseteq (A \cup B) \\ &\implies \{x\}, \{x, y\} \in \mathcal{P}(A \cup B) \\ &\implies \{\{x\}, \{x, y\}\} \subseteq \mathcal{P}(A \cup B) \\ &\implies (x, y) \in \mathcal{P}(\mathcal{P}(A \cup B)), \end{aligned}$$

so that we can take $C = \mathcal{P}(\mathcal{P}(A \cup B))$. ⊣

4.4. Ordered pairs. We now fix a specific definite operation (x, y) which satisfies 4.1 and 4.2, perhaps the Kuratowski pair defined in the proof of 4.3, perhaps some other: from now on we may forget the specific definition chosen, the only thing that counts is that the pair satisfies 4.1 and 4.2.

4.5. Exercise. *Let*

$$\text{Pair}(z) \iff_{\text{df}} (\exists x)(\exists y)[z = (x, y)], \quad (4.2)$$

$$\text{First}(z) =_{\text{df}} \begin{cases} \text{the unique } x \text{ such that } (\exists y)[z = (x, y)], & \text{if } \text{Pair}(z), \\ z, & \text{otherwise,} \end{cases} \quad (4.3)$$

$$\text{Second}(z) =_{\text{df}} \begin{cases} \text{the unique } y \text{ such that } (\exists x)[z = (x, y)], & \text{if } \text{Pair}(z), \\ z, & \text{otherwise.} \end{cases} \quad (4.4)$$

It follows that

$$\text{Pair}(z) \iff z = (\text{First}(z), \text{Second}(z)).$$

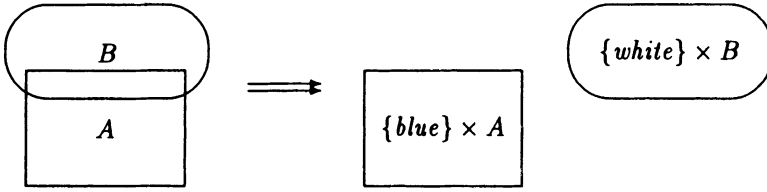


Figure 4.1. Constructing the disjoint union.

Using the ordered pair we can easily define triples, quadruples, etc. as well as the corresponding products, e.g.

$$(x, y, z) =_{\text{df}} (x, (y, z)), \quad (4.5)$$

$$(x, y, z, w) =_{\text{df}} (x, (y, z, w)) = (x, (y, (z, w))), \quad (4.6)$$

$$A \times B \times C =_{\text{df}} A \times (B \times C), \quad (4.7)$$

etc. By this definition, a tuple of length $n + 1$ is a pair with second member a tuple of length n .

4.6. Exercise. For all x, y, z, x', y', z' ,

$$(x, y, z) = (x', y', z') \iff x = x' \ \& \ y = y' \ \& \ z = z'.$$

4.7. Disjoint union. For each set A and fixed object *blue*, we can think of the set of pairs $\{\text{blue}\} \times A$ as a “blue copy” of A , the act of replacing each $a \in A$ by the pair (blue, a) being the set theoretic equivalent of painting a *blue*. We fix two such distinct “colors,”

$$\text{blue} =_{\text{df}} \emptyset, \quad \text{white} =_{\text{df}} \{\emptyset\}, \quad (4.8)$$

and we define the **disjoint union** of two sets by the formula

$$A \uplus B =_{\text{df}} (\{\text{blue}\} \times A) \cup (\{\text{white}\} \times B).$$

The notion is useful, and it should be clear that the specific identity of *blue* and *white* must be deliberately and instantly forgotten, all that matters is that $\text{blue} \neq \text{white}$.

4.8. Exercise. Show that for all sets A, B , $A \uplus \emptyset \subseteq A \uplus B$. Assuming that we use the Kuratowski pair to define products, give an example where the plausible inclusion $A \subseteq A \uplus B$ is not true.

Next we consider the notion of **relation** which permeates mathematics. Intuitively, a *binary relation* P between objects $x \in A$ and $y \in B$ is a

condition which is satisfied by some $x \in A$, $y \in B$ and fails for others. For example, the relation

$$xRy \iff_{\text{df}} x \text{ is a son of } y$$

is defined on $A = \{\text{men}\}$, $B = \{\text{women}\}$ and holds for x, y precisely if y has given birth to x . The obvious way to represent a binary relation in set theory is to identify it with its *extension*, the set of pairs which satisfy it.

4.9. Definition. A **binary relation** on the sets A, B is any subset R of the Cartesian product $A \times B$. We will use synonymously the notations

$$xRy \iff_{\text{df}} (x, y) \in R.$$

Obvious examples of binary relations are the **identity** and the relations of **membership** and **subsethood** restricted to some set A ,

$$\begin{aligned} x =_A y &\iff_{\text{df}} x \in A \ \& \ y \in A \ \& \ x = y, \\ x \in_A y &\iff_{\text{df}} x \in A \ \& \ y \in A \ \& \ x \in y, \\ X \subseteq_A Y &\iff_{\text{df}} X \subseteq Y \subseteq A, \end{aligned}$$

which by the definition are identified respectively with the sets

$$=_A =_{\text{df}} \{(x, y) \in A \times A \mid x = y\}, \quad (4.9)$$

$$\in_A =_{\text{df}} \{(x, y) \in A \times A \mid x \in y\}, \quad (4.10)$$

$$\subseteq_A =_{\text{df}} \{(X, Y) \in \mathcal{P}(A) \times \mathcal{P}(A) \mid X \subseteq Y\}. \quad (4.11)$$

4.10. Relations and definite conditions. The definite conditions $x = y$, $x \in y$ and $X \subseteq Y$ on the domain of all objects are not relations according to 4.9, in fact they are not even “coextensive” with sets of pairs, because their extensions are “too large.” This is important, the distinction between relations and definite conditions: briefly, *every relation determines a definite condition but (in general) the converse does not hold*. The precise situation is detailed in the next Exercise. In practice we will often refer to *the relation* $=$ on the set A , meaning (without possibility of confusion) the restriction $=_A$ as we just defined it.

4.11. Exercise. (1) For every binary relation $R \subseteq (A \times B)$, the condition

$$R^*(x, y) \iff_{\text{df}} (x, y) \in R$$

is *definite*. (Nothing to prove here, unless you want to practice applying 3.18.)

(2) For every binary definite condition P and any two sets A, B , the **restriction**

$$P_{A,B} =_{\text{df}} \{(x, y) \in A \times B \mid P(x, y)\}$$

of P to $A \times B$ is a binary relation.

Binary relations with both arguments ranging over the same set are especially important, and they are classified and studied according to the structural properties they may enjoy. Here are three such properties which come up often, in various combinations.

4.12. Definition. For each binary relation $P \subseteq (A \times A)$ on a set A we say that:

$$\begin{aligned} P \text{ is reflexive} &\iff_{\text{df}} (\forall x \in A)[xPx], \\ P \text{ is symmetric} &\iff_{\text{df}} (\forall x, y \in A)[xPy \implies yPx], \\ P \text{ is transitive} &\iff_{\text{df}} (\forall x, y, z \in A)[[xPy \ \& \ yPz] \implies xPz]. \end{aligned}$$

We call P an **equivalence relation** on A if it has all three of these properties. Equivalence relations are very useful and we will meet examples of them in practically every chapter of these Notes. They are often denoted by symbols like $\sim$, $\approx$, $\simeq$, so that their three characteristic properties take the form

$$\begin{aligned} x &\sim x, \\ x \sim y &\implies y \sim x, \\ x \sim y \ \& \ y \sim z &\implies x \sim z. \end{aligned}$$

4.13. Exercise. On each set A , the identity relation $\{(x, y) \mid x = y \in A\}$, the identically true relation $\{(x, y) \mid x, y \in A\}$ and for each $B \subseteq A$ the relation

$$x \sim_{A/B} y \iff_{\text{df}} x = y \vee [x \in B \ \& \ y \in B]$$

are all equivalence relations.

4.14. Proposition. Suppose $\sim$ is an equivalence relation on the set A , and for each $x \in A$ let

$$[x/\sim] = \{y \in A \mid x \sim y\} \tag{4.12}$$

be the **equivalence class**² of x . We denote the set of all these equivalence classes by

$$[A/\sim] = \{[x/\sim] \in \mathcal{P}(A) \mid x \in A\}. \tag{4.13}$$

It follows that for each $x \in A$, $[x/\sim] \neq \emptyset$, and

$$x \sim y \iff [x/\sim] = [y/\sim], \tag{4.14}$$

$$x \not\sim y \iff [x/\sim] \cap [y/\sim] = \emptyset. \tag{4.15}$$

²Each $[x/\sim]$ is obviously a set, a subset of A , and it would be more appropriately named the “equivalence set” of x , but the classical terminology goes way back and has been frozen.

Conversely, for each family $\mathcal{E}$ of non-empty and pairwise disjoint subsets of A such that $A = \bigcup \mathcal{E}$, the relation

$$x \sim y \iff_{\text{df}} (\exists X \in \mathcal{E})[x \in X \ \& \ y \in X]$$

is an equivalence relation on A and $\llbracket A/\sim \rrbracket = \mathcal{E}$.

Proof. Each $[x/\sim] \neq \emptyset$, since $x \in [x/\sim]$. By the transitivity and symmetry of $\sim$,

$$t \sim x \ \& \ x \sim y \implies t \sim y, \quad t \sim y \ \& \ x \sim y \implies t \sim x$$

so that

$$\begin{aligned} x \sim y &\implies (\forall t \in A)[t \sim x \iff t \sim y] \\ &\implies [x/\sim] = [y/\sim]. \end{aligned}$$

This implies immediately both (4.14) and (4.15). For the converse, the reflexivity and symmetry of $\sim$ are trivial. If $x \sim y$ and $y \sim z$, then there exist sets X, Y in $\mathcal{E}$ such that $x, y \in X$, $y, z \in Y$, so in particular $y \in X \cap Y$ and since the sets in $\mathcal{E}$ are pairwise disjoint, we have $X = Y$, so $x \sim z$. $\dashv$

4.15. Exercise. What are the equivalence classes of the equivalence relations in Exercise 4.13?

Following up the same idea, we identify each **ternary relation** R on the sets A, B, C with the set of triples which satisfy it, so that a ternary relation on A, B, C is simply an arbitrary subset of $A \times B \times C$. We will use synonymously the notations

$$R(x, y, z) \iff_{\text{df}} (x, y, z) \in R.$$

As with relations, we represent functions in set theory by identifying them with their “graphs.”

4.16. A function (or mapping or transformation) $f : A \rightarrow B$ with domain the set A and range the set B is any subset $f \subseteq (A \times B)$ which satisfies the condition

$$(\forall x \in A)(\exists! y \in B)[(x, y) \in f],$$

in more detail,

$$\begin{aligned} &(\forall x \in A)(\exists y \in B)[(x, y) \in f], \\ &(x, y) \in f \ \& \ (x, y') \in f \implies y = y'. \end{aligned}$$

For each $x \in A$ and $f : A \rightarrow B$, we will write, as usual,

$$\begin{aligned} f(x) &=_{\text{df}} \text{the unique } y \in B \text{ such that } (x, y) \in f \\ &= \text{the value of } f \text{ on the element } x. \end{aligned}$$

For any two sets A, B , we let

$$(A \rightarrow B) =_{\text{df}} \{f \subseteq A \times B \mid f : A \rightarrow B\} \quad (4.16)$$

be the set of all functions from A to B .

We will use all the familiar notations and abbreviations in connection with functions, e.g. sometimes writing the argument without the parentheses or as an index,

$$f(x) = fx = f_x.$$

The $\mapsto$ notation is also useful; for example, an **indexed family of sets** is a function

$$A = (i \mapsto A_i)_{i \in I} : I \rightarrow E$$

for some $I \neq \emptyset$ and some E , where each A_i is a set. We refer to I as the **index set** and we define the union and intersection of the family in the usual way,

$$\begin{aligned} \bigcup_{i \in I} A_i &=_{\text{df}} \{x \in \bigcup E \mid (\exists i \in I)[x \in A_i]\}, \\ \bigcap_{i \in I} A_i &=_{\text{df}} \{x \in \bigcup E \mid (\forall i \in I)[x \in A_i]\}. \end{aligned} \quad (4.17)$$

We can also define the **product** of an indexed family, the set of functions which select for each $i \in I$ one element from the value A_i ,

$$\prod_{i \in I} A_i =_{\text{df}} \{f : I \rightarrow \bigcup_{i \in I} A_i \mid (\forall i \in I)[f(i) \in A_i]\}. \quad (4.18)$$

Injectons, surjections, bijections (correspondences), images and pre-images of functions are defined exactly as in the Introduction. We will be using the notations:

$$\begin{aligned} (A \mapsto B) &=_{\text{df}} \{f \mid f : A \rightarrow B \text{ \& } f \text{ is an injection, one-to-one}\}, \\ (A \twoheadrightarrow B) &=_{\text{df}} \{f \mid f : A \rightarrow B \text{ \& } f \text{ is a surjection, onto } B\}, \\ (A \twoheadrightarrow B) &=_{\text{df}} (A \mapsto B) \cap (A \twoheadrightarrow B) \text{ (bijection, correspondence)}. \end{aligned}$$

We can use these to define equinumerosity and the size comparison condition with no reference to objects outside our theory,

$$A =_c B \iff_{\text{df}} (\exists f)[f : A \twoheadrightarrow B] \iff (A \twoheadrightarrow B) \neq \emptyset, \quad (4.19)$$

$$A \leq_c B \iff_{\text{df}} (\exists f)[f : A \mapsto B] \iff (A \mapsto B) \neq \emptyset. \quad (4.20)$$

4.17. Exercise. Prove from the axioms that $A =_c B \implies \mathcal{P}(A) =_c \mathcal{P}(B)$.

4.18. Exercise. Prove from the axioms that if $A =_c A'$ and $B =_c B'$, then

$$A \uplus B =_c A' \uplus B', \quad A \times B =_c A' \times B', \quad (A \rightarrow B) =_c (A' \rightarrow B').$$

For each $X \subseteq A$, the **restriction** $f \upharpoonright X$ of a function $f : A \rightarrow B$ is obtained by cutting f down so it is defined only on X ,

$$f \upharpoonright X =_{\text{df}} \{(x, y) \in f \mid x \in X\}. \quad (4.21)$$

It is also useful to notice that the basic condition of “functionhood”

$$\text{Function}(f) \iff_{\text{df}} (\exists A)(\exists B)[f \in (A \rightarrow B)] \quad (4.22)$$

is evidently definite. When we refer to a *function* f without identifying specific sets A, B such that $f : A \rightarrow B$, we will mean any set f which satisfies the condition $\text{Function}(f)$.

This identification in set theory of a function $f : A \rightarrow B$ with the set of pairs $\{(x, y) \in A \times B \mid f(x) = y\}$ has generated some controversy, because we have natural “operational” intuitions about the notion of function and by “function” we often mean a formula or a rule of computation. For example, the two functions on the reals

$$\begin{aligned} f(x, y) &=_{\text{df}} (x + y)^2, \\ g(x, y) &=_{\text{df}} x^2 + 2xy + y^2 \end{aligned}$$

are identified in set theory, although they are obviously different as computation rules. There is no problem with this if we keep clear in our minds that the “definition” 4.16 does not replace the intuitive notion of function but only represents it within *set theory*, faithfully for the uses to which we put this notion *within set theory*, e.g. to define equinumerosity.³

4.19. Cantor’s notion of cardinal numbers. Ironically, one of the most difficult, intuitive mathematical notions to represent faithfully in set theory is that of *cardinal number*, a most basic concept of the subject. Here is how Cantor introduced it in the same 1895 paper from which we quoted the “definition” of sets in the Introduction:

Every set A has a definite ‘power,’ which we will also call its ‘cardinal number’.

We will call by the name ‘power’ or ‘cardinal number’ of A the general concept, which by means of our active faculty of thought, arises from the set A when we make abstraction of its various elements x and of the order in which they are given.

We denote the result of this double act of abstraction, the cardinal number or power of A by $\overline{A}$. Since every element x ,

³Whether the intuitive notion of function-as-computation-rule can also be represented faithfully in set theory is an interesting problem, for which there does not exist yet a generally accepted solution.

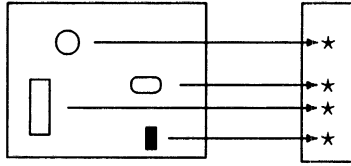


Figure 4.2. Cantor's construction of $|A|$ for a four-element set.

if we abstract from its nature becomes a 'unit,' the cardinal number $\overline{\overline{A}}$ is a definite set composed of units, and this number has existence in our minds as an intellectual image or projection of the given set A .

After some discussion, Cantor infers from this "definition" that cardinal numbers have the following two fundamental properties:

$$A =_c \overline{\overline{A}}, \quad (4.23)$$

$$A =_c B \iff \overline{\overline{A}} = \overline{\overline{B}}. \quad (4.24)$$

The first of these flows quite naturally from Cantor's conception: the process of abstraction which associates with each $x \in A$ a corresponding "unit" u_x evidently defines a correspondence $x \mapsto u_x$ between A and $\overline{\overline{A}}$. Cantor gives a brief argument for the second whose key phrase is that

$\overline{\overline{A}}$ grows, so to speak, out of A in such a way that from every element x of A a special unit of A arises.

To get (4.24) out of this we must assume that the "special units of A " depend only on "how many" members A has and not the nature of these members, which begs the question of cardinality, but there it is.

There is a third, more technical property of cardinal numbers, which Cantor uses routinely with no special mention to define and study operations which act on infinite families of sets: *for every family of sets* $\mathcal{E}$, $\{\overline{\overline{X}} \mid X \in \mathcal{E}\}$ is a set. Thus, however we understand Cantor's construction, it is quite clear what we must do to represent it faithfully in set theory. We substitute modern notation for Cantor's awkward double bar symbolism.

4.20. Problem of Cardinal Assignment: *to define an operation $|A|$ on the class of sets which satisfies*

$$A =_c |A|, \quad (4.25)$$

$$A =_c B \iff |A| = |B|, \quad (4.26)$$

$$\text{for each } \mathcal{E}, \{ |X| \mid X \in \mathcal{E} \} \text{ is a set.} \quad (4.27)$$

The problem is quite difficult and it was not solved until the twenties, by von Neumann, whose elegant construction uses both the Axioms of Choice and Replacement. We will present it in Chapter 12, as the culmination of a lot of work. In the meantime, notice that there are plenty of definite operations which satisfy (4.25) and (4.27), including the obvious $|A| = A!$ And as it turns out, these two properties suffice for the development of a very satisfactory theory of cardinality.

4.21. Cardinal numbers (1). A (weak) **cardinal assignment** is any definite operation $|A|$ which satisfies (4.25), $A =_c |A|$ and (4.27), i.e. for every family of sets $\mathcal{E}$, $\{|X| \mid X \in \mathcal{E}\}$ is a set. The **cardinal numbers** (relative to $|A|$) are its values,

$$\text{Card}(\kappa) \iff \kappa \in \text{Card} \iff_{\text{df}} (\exists A)[\kappa = |A|]. \quad (4.28)$$

A cardinal assignment $|A|$ is **strong** if in addition, for any two cardinal numbers κ, λ ,

$$\kappa =_c \lambda \iff \kappa = \lambda, \quad (4.29)$$

which is easily equivalent to (4.26).

We fix one, specific (possibly weak) cardinal assignment and we define the arithmetical operations on the cardinals as follows:

$$\begin{aligned} \kappa + \lambda &=_{\text{df}} |\kappa \uplus \lambda| &=_{\text{c}} \kappa \uplus \lambda, \\ \kappa \cdot \lambda &=_{\text{df}} |\kappa \times \lambda| &=_{\text{c}} \kappa \times \lambda, \\ \kappa^\lambda &=_{\text{df}} |(\lambda \rightarrow \kappa)| &=_{\text{c}} (\lambda \rightarrow \kappa). \end{aligned}$$

The infinitary operations are defined similarly:⁴

$$\begin{aligned} \sum_{i \in I} \kappa_i &=_{\text{df}} |\{(i, x) \in I \times \bigcup_{i \in I} \kappa_i \mid x \in \kappa_i\}|, \\ \prod_{i \in I} \kappa_i &=_{\text{df}} |\prod_{i \in I} \kappa_i|. \end{aligned}$$

The motivation is clear, e.g. the sum $\kappa + \lambda$ is the “number of elements” in the set we get by putting together disjoint copies of κ and λ .

4.22. Exercise. There is only one choice for $|\emptyset|$,

$$0 =_{\text{df}} |\emptyset| = \emptyset, \quad (4.30)$$

since only $|\emptyset| = \emptyset$ satisfies $\emptyset =_c |\emptyset|$. It is also convenient to set

$$1 =_{\text{df}} |\{0\}|, \quad 2 =_{\text{df}} |\{0, 1\}| \quad (4.31)$$

so we have handy names for the cardinals of a singleton and a doubleton.

⁴It is traditional to use the cap Greek Π to denote both the Cartesian product of sets and the cardinal operation of infinite product, and it does not really cause any confusion.

4.23. Exercise. For all cardinal numbers $\kappa_1 =_c \kappa_2$, $\lambda_1 =_c \lambda_2$,

$$\kappa_1 + \lambda_1 =_c \kappa_2 + \lambda_2, \quad \kappa_1 \cdot \lambda_1 =_c \kappa_2 \cdot \lambda_2, \quad \kappa_1^{\lambda_1} =_c \kappa_2^{\lambda_2}.$$

4.24. Cardinal arithmetic. It looks quite silly to develop the theory of a weak cardinal assignment which could be just the identity $|X| = X$, but the notation of cardinal numbers and the arithmetical operations on them is useful for expressing simply complex “equinumerosities.” Consider the formula

$$\kappa^{(\lambda+\mu)} =_c \kappa^\lambda \cdot \kappa^\mu. \quad (4.32)$$

It looks obvious, it is true by Problem **x4.15**, and it expresses exactly the same fact as

$$((\lambda \uplus \mu) \rightarrow \kappa) =_c (\lambda \rightarrow \kappa) \times (\mu \rightarrow \kappa), \quad (4.33)$$

more simply, or so some would say. More significantly, (1) the systematic development of formulas like (4.32) leads to a *cardinal arithmetic* which in the end suggests new (and useful) facts about equinumerosities by analogy with ordinary arithmetic, and (2) when we do construct von Neumann’s strong cardinal assignment, we will have already proved all the interesting facts about cardinals with $=_c$ in place of $=$: all we will need to do is remove in our minds the subscript $_c$ from facts we already understand!

The basic technique for proving identities of cardinal arithmetic is to use systematically (4.25) and its trivial consequence

$$A =_c B \iff |A| =_c |B|. \quad (4.34)$$

In connection with the arithmetical operations on cardinals, the replacement properties of the simple Exercise **4.23** are also very useful. To prove the associativity of cardinal addition, for example, we compute:

$$\begin{aligned} \kappa + (\lambda + \mu) &= \kappa \uplus (\lambda + \mu) && \text{by def.,} \\ &= \kappa \uplus (\lambda \uplus \mu) && \text{by def. and 4.18,} \\ &= (\kappa \uplus \lambda) \uplus \mu && \text{by a direct argument,} \\ &= (\kappa + \lambda) + \mu && \text{reversing the steps.} \end{aligned}$$

The mathematical essence of the proof is the alleged “direct argument,” which in this case is quite easy.

To see how the more technical condition (4.27) comes into play, consider the equation

$$|\bigcup_{i \in I} A_i| =_c \sum_{i \in I} |A_i|, \quad (4.35)$$

which should certainly be true when the sets in the family $(i \mapsto A_i)_{i \in I}$ are pairwise disjoint. To make sense of it, before trying to prove it, we must know that there is a function $(i \mapsto |A_i|)$, and that is exactly where (4.27) is used via the following.

4.25. Lemma. *For every indexed family of sets $A = (i \mapsto A_i)_{i \in I}$, there exists a function $f : I \rightarrow f[I]$ such that*

$$f(i) = |A_i| \quad (i \in I).$$

Proof. By (4.27) with

$$\mathcal{E} = \{A_i \mid i \in I\} = A[I],$$

there exists a set W which contains every $|A_i|$ for $i \in I$, and we can set

$$f =_{\text{df}} \{(i, w) \in I \times W \mid w = |A_i|\}. \quad \dashv$$

As it happens, equations like (4.35) cannot be proved without the Axiom of Choice, so we will have little need of (4.27) before Chapter 8.

4.26. Structured sets. A *topological space* is a set X of *points* endowed with a *topological structure*, which is determined by a collection $\mathcal{T}$ of subsets of X satisfying the following three properties:

1. $\emptyset, X \in \mathcal{T}$.
2. $A, B \in \mathcal{T} \implies A \cap B \in \mathcal{T}$.
3. For every family $\mathcal{E} \subseteq \mathcal{T}$ of sets in $\mathcal{T}$, the unionset $\bigcup \mathcal{E}$ is also in $\mathcal{T}$.

A family of sets $\mathcal{T}$ with these properties is called a **topology** on X , with **open sets** its members and **closed sets** the complements of open sets relative to X , i.e. all $X \setminus G$ with G open.

Notions like this of sets “endowed” with structure abound in mathematics: there are graphs, groups, vector spaces, sheaves, manifolds, partially ordered sets, etc. etc. In each of these cases we have a set X , typically called “the space,” and a complex of related objects which impose a structure on the space—functions, families of sets, other spaces with their own structure, etc. The pairing operation provides a simple and flexible way to model such notions faithfully in set theory.

A structured set is a pair

$$U = (A, \mathcal{S}), \tag{4.36}$$

where $A = \text{Field}(U)$ is a set, the *field* or *space* of U , and $\mathcal{S}$ is an arbitrary object, the *frame*⁵ of U .

⁵It would be more suggestive to call $\mathcal{S}$ *the structure* of the structured set $(A, \mathcal{S})$, but the word is heavily overloaded in logic and set theory and it is best to avoid attaching it to one more precise notion. Some people call “structures” what we have called “structured sets” here, at least when they are simple enough.

For example, a **topological space** is a structured set $(X, \mathcal{T})$, where the frame $\mathcal{T}$ is a topology on X , as above. A *group* is a structured set

$$U = (G, (e, \cdot)) \quad (4.37)$$

where $e \in G$ and $\cdot : G \times G \rightarrow G$ is a binary function, satisfying the *group axioms*, which do not concern us here. Notice that by the definition of triple (4.5), definition (4.37) is equivalent to

$$U = (G, e, \cdot). \quad (4.38)$$

It is quite common that the frame of a structured set is a tuple of objects, and then the structured set is also a tuple, with its field as the first element. We will meet numerous examples of this in the sequel.

Following usual mathematical practice, we will systematically confuse a structured set with its field when the frame is understood from the context or is not relevant. For example, we will refer to “the topological space X ” rather than “ $(X, \mathcal{T})$,” with “points” the members of X , “subsets” the subsets of X , etc. In the general case, the members of a structured set U are the members of $Field(U)$,

$$x \in U \iff_{\text{df}} x \in Field(U), \quad (4.39)$$

the subsets of U are the subsets of $Field(U)$, etc. Notice that the terminological convention (4.39) cannot possibly cause a misunderstanding: since we have (deliberately) not settled on a specific pairing operation—and have even left open the possibility that $(A, \mathcal{S})$ may be an atom (!)—the statement

$$x \in (A, \mathcal{S})$$

cannot possibly mean anything until we define it, and we just did this by (4.39).

Problems

The definition of ordered pair in the proof of 4.3 is due to the Polish set theorist and topologist Kuratowski. A few years before Kuratowski's construction, the American analyst Wiener had discovered the following, somewhat more complex but interesting solution of this problem.

x4.1. (Wiener) The properties 4.1 and 4.2 hold with the following definition of pair:

$$(x, y) =_{\text{df}} \{\{\emptyset, \{x\}\}, \{\{y\}\}\}.$$

x4.2. Prove from the axioms that for all sets A, B, C ,

$$((A \times B) \rightarrow C) =_c (A \rightarrow (B \rightarrow C)).$$

x4.3. Prove from the axioms the theorem of Cantor **2.21**, that for every set A , $A <_c \mathcal{P}(A)$. Which axioms do you need?

x4.4. For each function f , the **domain of definition** of f

$$\text{Domain}(f) =_{\text{df}} \{x \mid (\exists y)[(x, y) \in f]\}$$

and the **image** of f

$$\text{Image}(f) =_{\text{df}} \{y \mid (\exists x)[(x, y) \in f]\}$$

are sets, and for each set B ,

$$\text{Image}(f) \subseteq B \implies f : \text{Domain}(f) \rightarrow B.$$

As a consequence,

$$\text{Function}(f) \implies f : \text{Domain}(f) \rightarrow \text{Image}(f).$$

x4.5. A binary relation $\sim \subseteq (A \times A)$ is an equivalence relation on A if and only if there exists some set Q and a surjection

$$\pi : A \twoheadrightarrow Q \tag{4.40}$$

such that

$$x \sim y \iff \pi(x) = \pi(y). \tag{4.41}$$

When (4.40) and (4.41) hold, we call Q a **quotient** of A by $\sim$ and π a **determining surjection** of $\sim$. The proof of **4.14** yields the quotient $\llbracket A/\sim \rrbracket$ and the determining surjection $(x \mapsto [x/\sim])$, but in specific cases there exist other quotients which help us understand better the structure of the equivalence relation at hand.

x4.6. Suppose $\sim$ is an equivalence relation on A and $\pi : A \rightarrow A$ satisfies

$$x \sim y \implies \pi(x) = \pi(y) \in [x/\sim].$$

Prove that π is a determining surjection witnessing that its image $\pi[A] \subseteq A$ is a quotient of A by $\sim$.

x4.7. Fix an element $x_0 \in A$ in some set and define on the function space $(A \rightarrow B)$ the relation

$$f \sim g \iff_{\text{df}} f(x_0) = g(x_0).$$

Prove that $\sim$ is an equivalence relation and find a determining surjection $\pi : (A \rightarrow B) \rightarrow B$ which witnesses that B is a quotient of $(A \rightarrow B)$ by $\sim$.

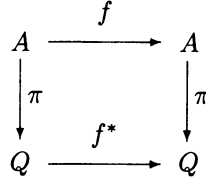


Figure 4.3.

x4.8. Let $x_0 \neq x_1$ be two distinct elements in A and find a determining surjection which witnesses that $(B \times B)$ is a quotient of $(A \rightarrow B)$ by the equivalence relation

$$f \sim g \iff_{\text{df}} f(x_0) = g(x_0) \ \& \ f(x_1) = g(x_1).$$

x4.9. Suppose $\sim$ is an equivalence relation on A and $f : A \rightarrow A$ is a function which **respects** $\sim$, i.e.

$$x \sim y \implies f(x) \sim f(y).$$

Let Q be any quotient of A by $\sim$. Prove that there exists a unique function $f^* : Q \rightarrow Q$ such that the diagram in Figure 4.3 *commutes*, i.e. $\pi f = f^* \pi$,

$$f^*(\pi x) = \pi(f(x)), \quad (x \in A),$$

where $\pi : A \rightarrow Q$ is a determining surjection.

x4.10. For all cardinal numbers, κ, λ, μ ,

$$\kappa + 0 =_c \kappa, \quad \kappa \cdot 0 =_c 0, \quad \kappa \cdot 1 =_c \kappa.$$

x4.11. For all cardinal numbers κ, λ, μ ,

$$\begin{aligned}
 \kappa + (\lambda + \mu) &= _c (\kappa + \lambda) + \mu, \\
 \kappa + \lambda &= _c \lambda + \kappa.
 \end{aligned}$$

x4.12. For all cardinal numbers κ, λ, μ ,

$$\begin{aligned}
 \kappa \cdot (\lambda \cdot \mu) &= _c (\kappa \cdot \lambda) \cdot \mu, \\
 \kappa \cdot \lambda &= _c \lambda \cdot \kappa, \\
 \kappa \cdot (\lambda + \mu) &= _c \kappa \cdot \lambda + \kappa \cdot \mu.
 \end{aligned}$$

x4.13. For all cardinals κ , $|\mathcal{P}(\kappa)| =_c 2^\kappa$.

x4.14. For all cardinal numbers κ, λ, μ ,

$$\kappa^0 =_c 1, \quad \kappa^1 =_c \kappa, \quad \kappa^2 =_c \kappa \cdot \kappa.$$

x4.15. For all cardinal numbers κ, λ, μ ,

$$\begin{aligned}(\kappa \cdot \lambda)^\mu &=_{\mathbf{c}} \kappa^\mu \cdot \lambda^\mu, \\ \kappa^{(\lambda+\mu)} &=_{\mathbf{c}} \kappa^\lambda \cdot \kappa^\mu, \\ (\kappa^\lambda)^\mu &=_{\mathbf{c}} \kappa^{\lambda \cdot \mu}.\end{aligned}$$

x4.16. For all cardinal numbers κ, λ, μ

$$\begin{aligned}\kappa \leq_{\mathbf{c}} \mu &\implies \kappa + \lambda \leq_{\mathbf{c}} \mu + \lambda, \\ \kappa \leq_{\mathbf{c}} \mu &\implies \kappa \cdot \lambda \leq_{\mathbf{c}} \mu \cdot \lambda, \\ \lambda \leq_{\mathbf{c}} \mu &\implies \kappa^\lambda \leq_{\mathbf{c}} \kappa^\mu, \\ \kappa \leq_{\mathbf{c}} \lambda &\implies \kappa^\mu \leq_{\mathbf{c}} \lambda^\mu.\end{aligned}$$

x4.17. For all A, B and all cardinals κ, λ ,

$$\prod_{i \in A} B = (A \rightarrow B), \quad \prod_{i \in \lambda} \kappa = \kappa^\lambda.$$

x4.18. Suppose $a \neq b$ are two distinct objects and κ_a, κ_b are cardinals, and prove that

$$\begin{aligned}\kappa_a + \kappa_b &=_{\mathbf{c}} \sum_{i \in \{a, b\}} \kappa_i, \\ \kappa_a \cdot \kappa_b &=_{\mathbf{c}} \prod_{i \in \{a, b\}} \kappa_i.\end{aligned}$$

x4.19. Prove that for all indexed families of cardinals,

$$\kappa \cdot \sum_{i \in I} \lambda_i =_{\mathbf{c}} \sum_{i \in I} \kappa \cdot \lambda_i.$$

x4.20. Show that $\kappa \cdot \lambda = 0 \iff \kappa = 0 \vee \lambda = 0$. Show also one of the directions of the equivalence

$$\prod_{i \in I} \kappa_i = 0 \iff (\exists i \in I)[\kappa_i = 0]. \quad (4.42)$$

(If you think you can show both directions of (4.42), think again and find your error, since one direction requires the Axiom of Choice.)

x4.21. The notion of equivalence according to Zermelo **3.27** coincides with equinumerosity, i.e.

$$A =_{\mathbf{c}} B \iff A \sim_Z B.$$

The definition **2.6** of infinite and finite sets refers to the set N of natural numbers and we cannot study these concepts axiomatically before we give a definition of N directly from the axioms. There is, however, another, simpler definition of these notions which we can give now and which we will later prove (with the Axiom of Choice) equivalent to **2.6**.

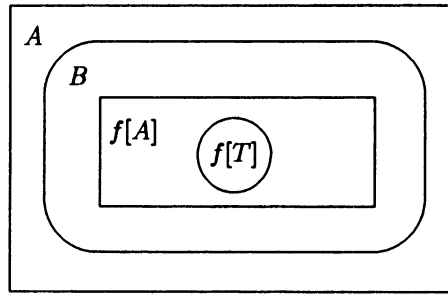


Figure 4.4. Zermelo's proof of the Schröder-Bernstein Theorem.

4.27. Definition. A set A is **infinite according to Dedekind** if there exists an injection

$$f : A \rightarrow B \subsetneq A$$

from A into a proper subset $B \subsetneq A$ (i.e. $B \neq A$). If A is not Dedekind-infinite, then it is **Dedekind-finite**.

x4.22. If A is Dedekind-infinite and $A =_c B$, then B is also Dedekind-infinite.

x4.23. If A is Dedekind-finite, then every subset of A is also Dedekind-finite.

x4.24. Every set I which satisfies the conditions

$$\emptyset \in I, \quad (\forall x)[x \in I \implies \{x\} \in I]$$

is Dedekind-infinite.

Most of the properties of Dedekind-finite sets require the Axiom of Choice for their proof. Here is one which does not, but it is quite difficult.

***x4.25.** If A, B are Dedekind-finite, then so is their union.

The classical proof of the Schröder-Bernstein Theorem **2.24** uses induction on the natural numbers and we cannot justify it now. In the next two problems we outline a very different proof (due to Zermelo), somewhat opaque in its motivation but elegant, short and in no way dependent on the natural numbers.

***x4.26.** If $A' \subseteq B \subseteq A$ and $A =_c A'$, then also $A =_c B$. **HINT:** Suppose $f : A \rightarrow A'$ is a correspondence which witnesses that $A =_c f[A] = A'$, and

$$Q = B \setminus f[A]$$

is the set of objects in B which are not in the image of A by f . We define the family of subsets of A

$$\mathcal{T} = \{X \mid Q \cup f[X] \subseteq X\}$$

and we first verify that its intersection is a member of it,

$$T =_{\text{df}} \bigcap \mathcal{T} \in \mathcal{T},$$

so that $Q \cup f[T] \subseteq T$. With a bit more work we can show that, in fact, $T = Q \cup f[T]$; this identity then implies that

$$B = T \cup (f[A] \setminus f[T]),$$

which completes the proof, since T and $(f[A] \setminus f[T])$ are disjoint sets and their union is (easily now) equinumerous with A .

***x4.27.** Use Problem ***x4.26** to give a proof of the Schröder-Bernstein Theorem from the axioms. HINT: If $f : A \rightarrow C$ and $g : C \rightarrow A$, then

$$A =_c gf[A] \subseteq g[C] \subseteq A, \quad g[C] =_c C.$$

Chapter 5

THE NATURAL NUMBERS

Our fundamental intuitive understanding of the natural numbers is that there is a (least) number 0, that every number n has a successor Sn , and that if we start with 0 and construct in sequence the successor of every number

$$0, S0 = 1, S1 = 2, S2 = 3, \dots$$

forever, then in time we will reach every natural number. In set theoretic terms we can capture this intuition by the following axiomatic characterization.

5.1. Definition. *A system of natural numbers is any structured set*

$$(N, 0, S) = (N, (0, S))$$

which satisfies the following conditions.

1. N is a set which contains the element 0, $0 \in N$.
2. S is a function on N , $S : N \rightarrow N$.
3. S is an injection, $Sn = Sm \implies n = m$.
4. For each $n \in N$, $Sn \neq 0$.
5. **Induction Principle.** For each $X \subseteq N$,

$$[0 \in X \ \& \ (\forall n \in N)[n \in X \implies Sn \in X]] \implies X = N.$$

These obvious properties of the natural numbers are called the **axioms of Peano** in honor of the Italian logician and mathematician who first proposed them as an axiomatic foundation of number theory. Most significant among them is the Induction Principle, whose typical application is illustrated in the proof of the next lemma.

5.2. Lemma. *In a system of natural numbers $(N, 0, S)$, every element $n \neq 0$ is a successor,*

$$n \neq 0 \implies (\exists m \in N)[n = Sm],$$

and for each n , $Sn \neq n$.

Proof. To prove the first assertion by the Induction Principle, it is enough to show that the set

$$X = \{n \in N \mid n = 0 \vee (\exists m \in N)[n = Sm]\}$$

satisfies the conditions

$$0 \in X, \quad (\forall n \in N)[n \in X \implies Sn \in X],$$

and both of these are obvious from the definition of X . In the same way, for the second assertion it is enough to verify that $S0 \neq 0$ (which holds because, in general, $Sn \neq 0$) and that $Sn \neq n \implies SSn \neq Sn$: this holds because S is one-to-one, so that $SSn = Sn \implies Sn = n$. $\dashv$

Number theory is one of the richest and most sophisticated fields of mathematics and it is by no means obvious that it can be developed on the basis of these five, simple properties; in fact, they do not suffice, one also needs to use set theory which (in its naive form) Peano took for granted, as part of “logic.” Here we will only show that the axioms imply the first, most basic properties of addition, multiplication and the ordering on the natural numbers, which is all we need. The proofs we will give, however, are characteristic samples of the use of the Peano axioms in the more advanced parts of the theory of numbers.

If number theory can be developed from the Peano axioms, then to give a faithful representation of the natural numbers in set theory, it is enough to prove from the axioms the following two theorems.

5.3. Existence Theorem for the Natural Numbers. *There exists at least one system of natural numbers $(N, 0, S)$.*

5.4. Uniqueness Theorem for the Natural Numbers. *For any two systems of natural numbers $(N_1, 0_1, S_1)$ and $(N_2, 0_2, S_2)$, there exists one (and only one) bijection*

$$\pi : N_1 \xrightarrow{\sim} N_2,$$

which satisfies the identities

$$\begin{aligned} \pi(0_1) &= 0_2, \\ \pi(S_1 n) &= S_2 \pi(n) \quad (n \in N_1). \end{aligned}$$

A bijection π which satisfies these identities is an **isomorphism** of the two systems $(N_1, 0_1, S_1)$ and $(N_2, 0_2, S_2)$, so that the theorem asserts that *any two systems of natural numbers are isomorphic*.

The Existence Theorem is very simple and we can prove it immediately.

5.5. Proof of the existence of natural numbers, 5.3. The Axiom of Infinity (VI) guarantees the existence of a set I such that

$$\emptyset \in I, \\ (\forall n)[n \in I \implies \{n\} \in I].$$

Using this I , first we define the family of sets

$$\mathcal{J} = \{X \subseteq I \mid \emptyset \in X \text{ \& } (\forall n)[n \in X \implies \{n\} \in X]\}$$

so that obviously $I \in \mathcal{J}$, and then we set

$$N = \bigcap \mathcal{J}, \quad 0 = \emptyset, \quad S = \{(n, m) \in N \times N \mid m = \{n\}\}.$$

To finish off the proof, it suffices to verify that this triple $(N, 0, S)$ is a system of natural numbers. To begin with, $N \in \mathcal{J}$, because $X \in \mathcal{J} \implies \emptyset \in X$ and hence $\emptyset \in \bigcap \mathcal{J} = N$, and by the same thinking,

$$n \in N \implies (\forall X \in \mathcal{J})[n \in X] \implies (\forall X \in \mathcal{J})[\{n\} \in X] \implies \{n\} \in N.$$

This implies immediately the first two of the Peano axioms, the next two hold because (in general, for all n, m) $\{n\} = \{m\} \implies n = m$ and $\{n\} \neq \emptyset$, and the Induction Principle follows directly from the definition of N as an intersection. $\dashv$

To prove the Uniqueness Theorem 5.4, we need the next fundamental result of axiomatic number theory.

5.6. Recursion Theorem. *Assume that $(N, 0, S)$ is a system of natural numbers, E is some set, $a \in E$, and $h : E \rightarrow E$ is some function: it follows that there exists exactly one function $f : N \rightarrow E$ which satisfies the identities*

$$f(0) = a, \\ f(Sn) = h(f(n)), \quad (n \in N).$$

The Recursion Theorem justifies the usual way by which we define functions on the natural numbers, **by recursion**¹ (or induction): to define $f : N \rightarrow E$, first we specify the value $f(0) = a$ and then we supply a function $h : E \rightarrow E$ which determines the value $f(Sn)$ of f at every successor Sn from the value $f(n)$ at its predecessor n , $f(Sn) = h(f(n))$. Our basic intuition about the natural numbers with which we started this chapter

¹The terms “recursion” and “induction” are often used synonymously in mathematics. We will follow the more recent convention which distinguishes **recursive definitions** from **inductive proofs**.

clearly justifies such definitions, so we should also be able to justify them on the basis of the axioms.

Before we establish the Recursion Theorem, let us use it in the next proof which is a typical example of the way it is applied.

5.7. Proof of the uniqueness of the natural numbers, 5.4. We assume that $(N_1, 0_1, S_1)$ and $(N_2, 0_2, S_2)$ are systems of natural numbers. By the Recursion Theorem on $(N_1, 0_1, S_1)$ with $E = N_2$, $a = 0_2$, $h = S_2$, there exists exactly one function $\pi : N_1 \rightarrow N_2$ which satisfies the identities

$$\begin{aligned}\pi(0_1) &= 0_2, \\ \pi(S_1 n) &= S_2 \pi(n) \quad (n \in N_1),\end{aligned}$$

and it suffices to verify that this π is a (one-to-one) correspondence.

(1) **π is a surjection**, $\pi : N_1 \rightarrow N_2$. Obviously $0_2 \in \pi[N_1]$ since $0_2 = \pi(0_1)$, and

$$\begin{aligned}m \in \pi[N_1] &\implies (\exists n \in N_1)[m = \pi(n)] \\ &\implies \pi(S_1 n) = S_2 \pi(n) = S_2 m \\ &\implies S_2 m \in \pi[N_1],\end{aligned}$$

so that by the Induction Principle on $(N_2, 0_2, S_2)$, $\pi[N_1] = N_2$.

(2) **π is an injection**, $\pi(n) = \pi(n') \implies n = n'$. It suffices to verify that the subset of N_1 ,

$$X = \{n \in N_1 \mid (\forall m \in N_1)[\pi(m) = \pi(n) \implies m = n]\},$$

satisfies the conditions

$$0_1 \in X, \quad n \in X \implies S_1 n \in X,$$

since together with the Induction Principle on $(N_1, 0_1, S_1)$, these imply $X = N_1$, which means that π is an injection. For the first condition,

$$\begin{aligned}m \neq 0_1 &\implies m = S_1 m' \text{ by Lemma 5.2} \\ &\implies \pi(m) = \pi(S_1 m') = S_2 \pi(m') \neq 0_2,\end{aligned}$$

so that $\pi(m) = \pi(0_1) = 0_2 \implies m = 0_1$ and $0_1 \in X$. For the second condition, it is enough to show that

$$n \in X \ \& \ \pi(m) = \pi(S_1 n) \implies m = S_1 n.$$

By the hypothesis

$$\pi(m) = \pi(S_1 n) = S_2 \pi(n) \neq 0_2,$$

which implies that $m \neq 0_1$, since $\pi(0_1) = 0_2$ and $0_1 \in X$. By Lemma 5.2 again, $m = S_1 m'$ for some $m' \in N_1$,

$$\pi(m) = \pi(S_1 m') = S_2 \pi(m')$$

and the hypothesis $\pi(m) = \pi(S_1 n)$ yields

$$S_2 \pi(m') = S_2 \pi(n),$$

which implies $\pi(m') = \pi(n)$. This, in turn, implies $m' = n$ because $n \in X$, so that $m = S_1 m' = S_1 n$, the required conclusion. $\dashv$

5.8. Proof of the Recursion Theorem. Assume the hypotheses and define first the set $\mathcal{A}$ of all *approximations* of the function which we want to construct:

$$\begin{aligned} p \in \mathcal{A} \quad &\Longleftrightarrow_{\text{df}} \text{Function}(p) \\ &\& \text{Domain}(p) \subseteq N \ \& \ \text{Image}(p) \subseteq E \\ &\& 0 \in \text{Domain}(p) \ \& \ p(0) = a \\ &\& (\forall n \in N)[Sn \in \text{Domain}(p) \\ &\implies n \in \text{Domain}(p) \ \& \ p(Sn) = h(p(n))]. \end{aligned} \tag{5.1}$$

We need to prove that there is exactly one function $f : N \rightarrow E$ (with domain of definition all of N) which belongs to $\mathcal{A}$.

Lemma. For all $p, q \in \mathcal{A}$ and $n \in N$,

$$n \in \text{Domain}(p) \cap \text{Domain}(q) \implies p(n) = q(n).$$

Proof. The set

$$X = \{n \in N \mid (\forall p, q \in \mathcal{A}) n \in \text{Domain}(p) \cap \text{Domain}(q) \implies p(n) = q(n)\}$$

clearly contains 0, since every $p \in \mathcal{A}$ satisfies $p(0) = a$. If

$$n \in X \ \& \ p \in \mathcal{A} \ \& \ q \in \mathcal{A} \ \& \ Sn \in \text{Domain}(p) \cap \text{Domain}(q),$$

then

$$\begin{aligned} p(Sn) &= h(p(n)) \quad \text{because } p \in \mathcal{A}, \\ &= h(q(n)) \quad \text{because } p(n) = q(n), \\ &= q(Sn) \quad \text{because } q \in \mathcal{A}, \end{aligned}$$

so that $n \in X \implies Sn \in X$, and hence, by the Induction Principle, $X = N$ and the Lemma is true.

The Lemma implies immediately that *at most one* function $f : N \rightarrow E$ belongs to $\mathcal{A}$, so to complete the proof of the theorem we need only show that *at least one* such f exists. This is the union

$$f = \bigcup \mathcal{A} = \{(n, w) \mid (\exists p \in \mathcal{A})[n \in \text{Domain}(p) \ \& \ p(n) = w]\},$$

which is a function, because

$$\begin{aligned} (n, w) \in f \ \& \ (n, w') \in f &\implies (\exists p, q \in \mathcal{A})[(n, w) \in p \ \& \ (n, w') \in q] \\ &\implies w = w' \text{ from the Lemma,} \end{aligned}$$

and then the definition of $\mathcal{A}$ and a similar calculation shows that $f \in \mathcal{A}$. The only thing left is to verify that $\text{Domain}(f) = N$, and for that we will use once more the Induction Principle. To begin with, $0 \in \text{Domain}(f)$, since $0 \in \text{Domain}(p)$ for every $p \in \mathcal{A}$. If $n \in \text{Domain}(f)$, then there exists some function $p \in \mathcal{A}$ with $n \in \text{Domain}(p)$, and hence (easily)

$$q = p \cup \{(Sn, h(p(n)))\} \in \mathcal{A},$$

so that $Sn \in \text{Domain}(q) \subseteq \text{Domain}(f)$. -1

5.9. The Natural Numbers. We now fix a specific system $(N, 0, S)$ of natural numbers whose members we will henceforth call **numbers** or **integers**. Following Cantor, we denote the cardinal number of N by the first Hebrew letter,

$$\aleph_0 =_{\text{df}} |N|. \tag{5.2}$$

Later we will meet its followers $\aleph_1, \aleph_2$, etc. Functions $a : N \rightarrow A$ with domain N are called (infinite) **sequences** and we often write their argument as a subscript,

$$a_n = a(n) \quad (n \in N, a : N \rightarrow A).$$

An obvious choice for N would be the system which we constructed in the proof of the Existence Theorem 5.3, where $0 = \emptyset$,

$$N = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\}$$

and $Sn = \{n\}$. Another choice which some would prefer on philosophical grounds is to assert that there exists, in fact, a set

$$N = \{0, 1, 2, \dots\}$$

of the “true natural numbers,” which are not sets, and the successor function S is nothing like the artificial $(n \mapsto \{n\})$, but it is the natural function which associates with each number n “the next number” Sn . Zermelo’s theory allows such non-sets (like the “true numbers”) as atoms and requires only one thing: that the system of natural numbers satisfies the Peano axioms, something which every serious person will surely grant. As far as the mathematical theory of numbers and sets is concerned, these two (and all other) choices of the objects we will call *numbers* are equivalent, since we will base all our proofs on the Peano axioms alone.

The Recursion Theorem is easier to apply in the following form.

5.10. Corollary. Recursion with parameters. *For any two sets Y , E and functions*

$$g : Y \rightarrow E, \quad h : E \times Y \rightarrow E,$$

there exists exactly one function $f : N \times Y \rightarrow E$ which satisfies the identities

$$\begin{aligned} f(0, y) &= g(y) & (y \in Y), \\ f(n+1, y) &= h(f(n, y), y) & (y \in Y, n \in N). \end{aligned}$$

Proof. For each $y \in Y$, we define the function $h_y : E \rightarrow E \times N$ by the formula

$$h_y(w) = h(w, y),$$

and by the Recursion Theorem we know that there exists *exactly one* function

$$f_y : N \rightarrow E \times N$$

which satisfies the identity

$$\begin{aligned} f_y(0) &= g(y), \\ f_y(n+1) &= h_y(f(n)) = h(f(n), y). \end{aligned}$$

It follows immediately that the function $f : N \times Y \rightarrow E$ defined by the formula

$$f(n, y) =_{\text{df}} f_y(n) \quad (y \in Y, n \in N)$$

satisfies the conclusion of the Corollary. $\dashv$

5.11. Addition and multiplication. *The addition function on the natural numbers is defined by the recursion*

$$\begin{aligned} n + 0 &= n, \\ n + (Sm) &= S(n + m), \end{aligned} \tag{5.3}$$

and multiplication is defined next, using addition, by the recursion

$$\begin{aligned} n \cdot 0 &= 0, \\ n \cdot Sm &= (n \cdot m) + n. \end{aligned} \tag{5.4}$$

In more detail, we know from **5.10** that there exists exactly one function $f : N \times N \rightarrow N$ which satisfies the identities

$$\begin{aligned} f(0, n) &= g(n), \\ f(Sm, n) &= h(f(m, n), n), \end{aligned}$$

where the functions g and h have been given as sets of pairs,

$$\begin{aligned} g &= \{(n, n) \in N \times N \mid n \in N\}, \\ h &= \{((z, n), w) \in (N \times N) \times N \mid w = Sz\}, \end{aligned}$$

and we define addition by

$$n + m = f(m, n),$$

i.e. $+$ $= \{((n, m), w) \mid ((m, n), w) \in f\}$. Such scholastic details do not enhance understanding (rather the opposite) and we will avoid them in the future.

5.12. Theorem. *Addition is associative, i.e. it satisfies the identity*

$$(n + m) + k = n + (m + k) \quad (5.5)$$

Proof. First for $k = 0$,

$$(n + m) + 0 = n + m = n + (m + 0),$$

using twice the identity $w + 0 = w$ directly from the definition of addition. Inductively, assuming that for some k

$$(n + m) + k = n + (m + k), \quad (5.6)$$

we compute:

$$\begin{aligned} (n + m) + Sk &= S((n + m) + k) \\ &= S(n + (m + k)) \quad \text{by (5.6)} \\ &= n + S(m + k) \\ &= n + (m + Sk) \end{aligned}$$

where the steps we did not justify follow from the definition of addition. $\dashv$

The commutativity of addition is not quite so simple and requires two lemmas.

5.13. Lemma. *For every natural number n , $0 + n = n$.*

Proof. By induction, $0 + 0 = 0$ follows from the definition, and if $0 + n = n$, then $0 + Sn = S(0 + n) = Sn$. $\dashv$

5.14. Lemma. *For all n, m , $n + Sm = Sn + m$.*

Proof. By induction on m , first for $m = 0$, immediately from the definition:

$$n + S0 = S(n + 0) = Sn = Sn + 0.$$

At the induction step, we assume that for some m

$$n + Sm = Sn + m \quad (5.7)$$

and we must show that

$$n + SSm = Sn + Sm.$$

Compute:

$$\begin{aligned} n + SSm &= S(n + Sm) && \text{by the definition} \\ &= S(Sn + m) && \text{from (5.7)} \\ &= Sn + Sm && \text{by the definition.} \end{aligned} \quad \dashv$$

5.15. Theorem. *Addition is a commutative function, i.e. it satisfies the identity*

$$n + m = m + n.$$

Proof is by induction on m , the basis being immediate from Lemma 5.13. At the induction step, we assume that for some specific m

$$n + m = m + n \quad (5.8)$$

and we compute:

$$\begin{aligned} n + Sm &= S(n + m) && \text{by the definition} \\ &= S(m + n) && \text{from (5.8)} \\ &= m + Sn && \text{by the definition} \\ &= Sm + n && \text{by Lemma 5.14.} \end{aligned} \quad \dashv$$

5.16. Exercise. *For every natural number n , the function $(s \mapsto n + s)$ is one-to-one, so that $n + s = n + t \implies s = t$, and in particular*

$$n + s = n \implies s = 0.$$

5.17. Definition. *A binary relation $\leq$ on a set P is a **partial ordering** if it is reflexive, transitive and antisymmetric, i.e. for all x, y ,*

$$\begin{aligned} x &\leq x, && \text{reflexivity} \\ x \leq y \ \&\ y \leq z \implies x \leq z, && \text{transitivity} \\ x \leq y \ \&\ y \leq x \implies x = y, && \text{antisymmetry.} \end{aligned}$$

In connection with partial orderings we will also use the notation

$$x < y \iff x \leq y \ \& \ x \neq y.$$

The partial ordering $\leq$ is **total**, **linear**, or simply an **ordering**, if, in addition, any two elements of P are **comparable** in $\leq$, i.e.

$$(\forall x, y \in P)[x \leq y \vee y \leq x],$$

or equivalently

$$(\forall x, y \in P)[x < y \vee x = y \vee y < x].$$

5.18. Definition. *The binary relation $\leq$ on P is a **wellordering** of P , if it is a total ordering of P and, in addition, every non-empty subset of P has a least element,*

$$(\forall X \subseteq P)[X \neq \emptyset \implies (\exists x \in X)(\forall y \in X)[x \leq y]].$$

Correct English would have us call these “good orderings,” and in fact this is what they are called in every other language, but the awkward “wellordering” has been established so firmly that it is hopeless to try and change it.

5.19. Definition. *The order relation $\leq$ on the natural numbers is defined by the equivalence*

$$n \leq m \iff_{\text{df}} (\exists s)[n + s = m].$$

The most basic property of $\leq$ is:

5.20. Lemma. *For all natural numbers n, m ,*

$$n \leq Sm \iff n \leq m \vee n = Sm.$$

Proof. If $n \leq Sm$, then there is some t such that $n + t = Sm$ by the definition, and we consider two cases. CASE (1), $t = 0$. Now $n + 0 = Sm$, hence $n = Sm$. CASE (2), $n + t = Sm$ for some $t \neq 0$. Now, by (5.2), $t = Ss$ for some s , so that $n + Ss = Sm$, hence $S(n + s) = Sm$, hence $n + s = m$ because S is an injection and hence $n \leq m$. The converse direction of the Lemma is easier. $\dashv$

5.21. Theorem. *The relation $\leq$ on the natural numbers is a wellordering.*

Proof. Reflexivity is immediate from $n + 0 = n$ and transitivity holds because $n + s = m$ & $m + t = k \implies n + (s + t) = k$. To prove antisymmetry, notice that if $n + s = m$ and $m + t = n$, then $n + (s + t) = n$ and by Exercise 5.16 we have $s + t = 0$; this implies $t = 0$ (otherwise $s + t$ is a successor) and hence $m = n$.

Proof of linearity. We show that $(\forall n)[n \leq m \vee m \leq n]$, by induction on m . Notice first that for every n , $n \leq Sn$, because $n + S0 = Sn$.

BASIS. For every n , $0 + n = n$ and hence $0 \leq n$.

INDUCTION STEP. We assume the induction hypothesis

$$(\forall n)[n \leq m \vee m \leq n]$$

and show that for each n , $n \leq Sm \vee Sm \leq n$. The induction hypotheses naturally splits the proof up in two cases. If $n \leq m$, then $n \leq Sm$ because $m \leq Sm$ and $\leq$ is transitive. If $m \leq n$, then for some t , $m + t = n$, and again we have two cases: if $t = 0$, then $n = m \leq Sm$, and if $t \neq 0$, then $t = Ss$ for some s , so $m + Ss = n$ and from (5.14) $Sm + s = n$, hence $Sm \leq n$.

Proof of the wellordering property. Towards a contradiction, suppose that X is non-empty but has no least element and let

$$Y = \{n \in N \mid (\forall m \leq n)[m \notin X]\},$$

so that obviously

$$Y \cap X = \emptyset. \quad (5.9)$$

It is enough to show that $0 \in Y$ and $n \in Y \implies Sn \in Y$, because then $Y = N$ by the Induction Principle and hence $X = \emptyset$ by (5.9), which is a contradiction.

BASIS. $0 \in Y$. Since 0 is the least number, we must have $0 \notin X$ (otherwise X would have a least member) and also $m \leq 0 \implies m = 0 \implies m \notin X$, so $0 \in Y$.

INDUCTION STEP. The induction hypothesis $n \in Y$ and the definition of Y imply that $(\forall m \leq n)m \notin X$, and then we know from Lemma 5.20 that $m \leq Sn \implies m \leq n \vee m = Sn$. Hence to verify that $Sn \in Y$, it is enough to show that $Sn \notin X$. But if Sn were a member of X , then it would be the least member of X since

$$\begin{aligned} m < Sn &\iff m \leq n \quad \text{by (5.20)} \\ &\implies m \notin X \quad \text{by the ind. hyp.} \end{aligned}$$

This shows that $\leq$ has the wellordering property and completes the proof of the theorem. $\dashv$

About wellorderings, in general, we will say a lot in Chapter 7. In the special case of the natural numbers, the fact that N is well ordered by $\leq$ is another manifestation of the Induction Principle.

Before we begin studying the applications of recursion to the theory of finite and countable sets, we should recall the warning issued in 3.23: some of them require the Axiom of Choice and we will not be able to justify them axiomatically until we add that axiom to our system in Chapter 8.

Most, however, can be established by judicious applications of the general method of proof which can be symbolized by the coupling

recursive definition — inductive proof.

First we repeat the definitions of Chapter 2, with the axiomatic notions now at our disposal.

5.22. Definition. *For any two natural numbers $n \leq m$, the (half-open) interval from n to m is the set*

$$[n, m) =_{\text{df}} \{k \in N \mid n \leq k \ \& \ k < m\}.$$

5.23. Exercise. *For each n , $[n, n) = \emptyset$, and for all $n \leq m$,*

$$[n, Sm) = [n, m) \cup \{m\}.$$

5.24. Definition. *A set A is **finite** if there exists some natural number n such that $A =_c [0, n)$, **infinite** if it is not finite and **countable** if it is finite or equinumerous with N . The **finite cardinals** are the cardinal numbers of finite sets.*

The next crucial property of finite sets is the first, basic result in the field of *combinatorics*.

5.25. Pigeonhole Principle. *Every injection $f : A \rightarrow A$ on a finite set into itself is also a surjection, i.e. $f[A] = A$.*

Proof. It is enough to prove that for each natural number n and each g ,

$$g : [0, m) \rightarrow [0, m) \implies g[[0, m)] = [0, m), \quad (5.10)$$

for the following reason. If $f : A \rightarrow A$ is an injection and $\pi : A \rightarrow [0, m)$ witnesses that A is finite, we define $g : [0, m) \rightarrow [0, m)$ by the equation

$$g(i) = \pi(f(\pi^{-1}(i))) \quad (i < m),$$

so that (easily)

$$f(x) = \pi^{-1}(g(\pi(x))) \quad (x \in A). \quad (5.11)$$

Now g is an injection, as a composition of injections, so that from (5.10) it is a bijection; but then f is also a bijection, because it is a composition of bijections, by (5.11).

The proof of (5.10) is (naturally) by induction on m . It is important to note at the outset that what we will show is the general assertion

$$(\forall h) [h : [0, m) \rightarrow [0, m) \implies h[[0, m)] = [0, m)], \quad (5.12)$$

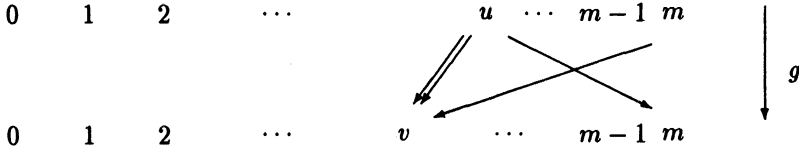


Figure 5.1. CASE 3 in the Pigeonhole Principle proof.

because in the verification of the induction step for some g we will need the induction hypothesis for various other h 's.

BASIS. (5.10) is trivial when $m = 0, 1$, because only one function $g : [0, m) \rightarrow [0, m)$ exists in these cases and it is a bijection.

INDUCTION STEP. We assume (5.12) for some $m > 1$ and we proceed to prove that every injection

$$g : [0, Sm) \rightarrow [0, Sm)$$

is a surjection. From Exercise 5.23 we know that

$$[0, Sm) = [0, m) \cup \{m\},$$

and the proof naturally splits into three cases.

CASE (1). $m \notin \text{Image}(g)$. Consider the *restriction* h of g to the interval $[0, m)$, which is defined by

$$h(i) = g(i) \quad (i < m),$$

i.e. as a set of pairs, $h = g \setminus \{(m, g(m))\}$. This takes all its values in $[0, m)$ and it is certainly an injection, so the induction hypothesis holds for it and hence $h[[0, m)] = [0, m)$. This means that $g[[0, m)] = [0, m)$, which is absurd, because the Case Hypothesis implies $g(m) < m$ so that the value $g(m)$ is taken on twice and g is not an injection.

CASE (2). $g(m) = m$. By the same reasoning, the restriction h is a bijection $h : [0, m) \rightarrow [0, m)$, and hence (trivially now) g is also a bijection.

CASE (3). There exist numbers $u, v < m$ such that

$$g(u) = m, \quad g(m) = v.$$

In this most interesting case, we define the function $h' : [0, m) \rightarrow [0, m)$ by the formula

$$h'(i) = \begin{cases} g(i), & \text{if } i < m \text{ \& } i \neq u, \\ v, & \text{if } i = u. \end{cases}$$

Now h' is an injection, because it agrees with g at all arguments except u , where it takes the value v ; and $v \neq g(j)$, for every $j < m$, because $g(m) = v$ and g is an injection. The induction hypothesis applied to h' implies that $h'[[0, m)] = [0, m)$, and using this (easily), $g[[0, Sm)] = [0, Sm)$. $\dashv$

As a first application we can give a rigorous proof of the following “obvious” result.

5.26. Corollary. *The set N of natural numbers is infinite.*

Proof. The function $(n \mapsto Sn)$ is a non-trivial injection of N into N . $\dashv$

It follows that “infinite, countable” means precisely “equinumerous with N ,” in accordance with our basic intuitions: a set A is countable, infinite just when $|A| =_c \aleph_0$.

5.27. Corollary. *For each finite set A , there exists exactly one natural number n such that $A =_c [0, n]$. We let*

$$\#(A) =_{\text{df}} \text{the unique } n \in N [A =_c [0, n]] \quad (5.13)$$

and we naturally call $\#(A)$ the number of elements of A .

Proof. If $A =_c [0, n]$ and $A =_c [0, m]$ with $n < m$, then $[0, n] =_c [0, m]$ and the correspondence $\pi : [0, m] \rightarrow [0, n]$ contradicts the Pigeonhole Principle, since $[0, n]$ is a proper subset of $[0, m]$. $\dashv$

From this point on we can proceed to prove all the basic properties of finite sets by induction on the number of elements in them, which is essentially their cardinal number. The method is illustrated in the problems.

The proof of the Schröder-Bernstein Theorem 2.24 used another variant of definition by recursion.

5.28. Simultaneous Recursion Theorem. *For each two sets E_1, E_2 , elements $a_1 \in E_1, a_2 \in E_2$ and functions $h_1 : E_1 \times E_2 \rightarrow E_1, h_2 : E_1 \times E_2 \rightarrow E_2$, there exist unique functions*

$$f_1 : N \rightarrow E_1, \quad f_2 : N \rightarrow E_2$$

which satisfy the identities

$$\begin{aligned} f_1(0) &= a_1, & f_2(0) &= a_2, \\ f_1(n+1) &= h_1(f_1(n), f_2(n)), & f_2(n+1) &= h_2(f_1(n), f_2(n)). \end{aligned}$$

Proof. Apply the Recursion Theorem 5.6 to $E = E_1 \times E_2$, $a = (a_1, a_2)$ and

$$h(w_1, w_2) = (h_1(w_1, w_2), h_2(w_1, w_2))$$

to get a function $f : N \rightarrow E_1 \times E_2$ and then set

$$f_1(n) = \text{First}(f(n)), \quad f_2(n) = \text{Second}(f(n)),$$

using the component functions of Exercise 4.5. $\dashv$

The functions $(n \mapsto A_n)$ and $(n \mapsto B_n)$ in the proof of **2.24** are defined by simultaneous recursion with $E_1 = \mathcal{P}(A)$ and $E_2 = \mathcal{P}(B)$ and the rest of the argument is elementary and can be based on the axioms. (Check it!)

5.29. Strings. In Chapter 2 we used the n -fold Cartesian product A^n to represent sequences of length n from a given set A . This is not convenient when we wish to study the set of all finite sequences from A , and it is better to represent these using functions on initial segments of N . For each set A , we define the set of **finite sequences**, **words** or **strings** from A by

$$\begin{aligned} A^{(n)} &=_{\text{df}} \{u \subseteq N \times A \mid \text{Function}(u) \ \& \ \text{Domain}(u) = [0, n)\}, \\ A^* &=_{\text{df}} \bigcup_{n=0}^{\infty} A^{(n)}, \end{aligned} \quad (5.14)$$

and we let

$$lh(u) =_{\text{df}} \max\{i \mid i = 0 \vee i - 1 \in \text{Domain}(u)\} \quad (u \in A^*), \quad (5.15)$$

be the **length** of the string u , so that $lh(u) = 0$ exactly when $u = \emptyset$ is the empty string. We also let

$$u \sqsubseteq v \iff_{\text{df}} u \subseteq v \quad (u, v \in A^*), \quad (5.16)$$

and we call u an **initial segment** of v if $u \sqsubseteq v$. If $a_0, \dots, a_{n-1}$ are elements of A , we let

$$\langle a_0, \dots, a_{n-1} \rangle =_{\text{df}} \{(0, a_0), \dots, (n-1, a_{n-1})\} \quad (5.17)$$

be the sequence of these objects and, in particular (with $n = 0, 1$),

$$\langle \rangle = \emptyset, \quad \langle a \rangle =_{\text{df}} \{(0, a)\}. \quad (5.18)$$

For any two strings u, v , the string

$$u \star v = \langle u(0), \dots, u(lh(u) - 1), v(0), \dots, v(lh(v) - 1) \rangle \quad (5.19)$$

is the **concatenation** of the strings u and v . For each $f : N \rightarrow A$ and each natural number n ,

$$\bar{f}(n) =_{\text{df}} f \upharpoonright [0, n) = \{(i, f(i)) \mid i < n\} \quad (5.20)$$

is the **restriction** of f to the initial segment $[0, n)$ of N . For example,

$$\bar{f}(0) = \emptyset, \quad \bar{f}(1) = \{(0, f(0))\}, \dots,$$

and we can recover f from $\bar{f}$, since

$$i < n \implies f(i) = \bar{f}(n)(i).$$

3	9	10	11	
2	4	5	8	
1	1	3	7	
0	0	2	6	12
	0	1	2	3

Figure 5.2. A pairing of $N \times N$ with N .

5.30. Definition. For each cardinal number κ and each $n \in N$, we set

$$\kappa^n =_{\text{df}} |\kappa^{(n)}|.$$

5.31. Proposition. For each countably infinite set A and each $n > 0$,

$$A =_c A \times A =_c A^{(n)} =_c A^*.$$

As equations of cardinal arithmetic, these read:

$$\aleph_0 =_c \aleph_0 \cdot \aleph_0 =_c \aleph_0^n =_c |\aleph_0^*|. \quad (5.21)$$

Proof. The inequalities from left to right are trivial, so by the Schröder-Bernstein Theorem it is enough to show $N^* \leq_c N$. We need to start with some injection

$$\rho : N \times N \rightarrow N, \quad (5.22)$$

suppose we have one. Using it, we define by recursion an injection $\pi_n : N^{(n+1)} \rightarrow N$, for each n , so that

$$\begin{aligned} \pi_0(u) &= u(0), \\ \pi_{n+1}(u) &= \rho(\pi_n(u \upharpoonright [0, n+1)), u(n+1)); \end{aligned}$$

in full detail (for the last time), this comes from the Recursion Theorem, by setting

$$\begin{aligned} \pi_0 &= \{(u, w) \mid u \in N^{(1)}, (0, w) \in u\}, \\ \pi_{n+1} &= \{(u, w) \mid u \in N^{(n+2)}, w = \rho(\pi_n(u \upharpoonright [0, n+1]), u(n+1))\}. \end{aligned}$$

Finally, the function

$$\pi(u) = (lh(u) - 1, \pi_{lh(u)-1}(u))$$

proves that $\bigcup_{n=0}^{\infty} N^{(n+1)} \leq_c N$, from which the full result follows immediately by using ρ once more. As far as choosing a ρ to start with, everyone has their favorite way of coding pairs and Cantor's illustrated in Figure 2.2 will certainly do. Here is another one, due to Gödel and pictured in Figure 5.2:

$$\rho(m, n) = \begin{cases} (m+1)^2 - 1, & \text{if } m = n, \\ n^2 + m, & \text{if } m < n, \\ m^2 + m + n, & \text{if } n < m. \end{cases}$$

The proof that it actually works is fun. $\dashv$

5.32. The continuum. The classical notation for the cardinal of $\mathcal{P}(N)$ is

$$\mathfrak{c} =_{\text{df}} |\mathcal{P}(N)| =_c 2^{\aleph_0}. \quad (5.23)$$

The elementary facts about $\mathfrak{c}$ are easy to establish, using the properties of $\aleph_0$ in (5.21) and elementary cardinal arithmetic. For example,

$$\mathfrak{c} \cdot \mathfrak{c} =_c 2^{\aleph_0} \cdot 2^{\aleph_0} =_c 2^{\aleph_0 + \aleph_0} =_c 2^{\aleph_0} =_c \mathfrak{c}.$$

The Schröder-Bernstein Theorem is also very useful, for example

$$\mathfrak{c} =_c 2^{\aleph_0} \leq_c \aleph_0^{\aleph_0} \leq_c \mathfrak{c}^{\aleph_0} =_c (2^{\aleph_0})^{\aleph_0} =_c 2^{\aleph_0 \cdot \aleph_0} =_c 2^{\aleph_0} =_c \mathfrak{c},$$

which by Schröder-Bernstein implies that

$$\mathfrak{c} =_c \aleph_0^{\aleph_0} =_c \mathfrak{c}^{\aleph_0}.$$

Some of the problems ask for computations of this type. On the other hand, the equinumerosity $\mathcal{R} =_c \mathcal{P}(N)$ will follow from the axioms easily once we have defined the reals $\mathcal{R}$ in Appendix A, so the Continuum Hypothesis is equivalent to the proposition

$$(\mathbf{CH}) \quad (\forall \kappa \leq_c \mathfrak{c}) [\kappa \leq_c \aleph_0 \vee \kappa =_c \mathfrak{c}].$$

We will discuss **CH** in Chapter 10, it is not that easy to resolve.

Problems

x5.1. Multiplication on the natural numbers is associative.

x5.2. Multiplication on the natural numbers is commutative.

x5.3. Exponentiation on the natural numbers is defined by the following recursion on m :

$$\begin{aligned} n^0 &= 1, \\ n^{Sm} &= n^m \cdot n. \end{aligned}$$

Show that it satisfies the following identities (for $n \neq 0$):

$$\begin{aligned} n^{(m+k)} &= n^m \cdot n^k, \\ n^{(m \cdot k)} &= (n^m)^k. \end{aligned}$$

x5.4. Suppose $(N_1, 0_1, S_1)$ and $(N_2, 0_2, S_2)$ are systems of natural numbers, $+_1, \cdot_1, +_2, \cdot_2$ are the functions of addition and multiplication in these systems, and $\pi : N_1 \rightarrow N_2$ is the “canonical” isomorphism between them according to Theorem 5.4. Show that π is an isomorphism with respect to addition and multiplication also, i.e. for all $n, m \in N_1$,

$$\pi(n +_1 m) = \pi(n) +_2 \pi(m), \quad \pi(n \cdot_1 m) = \pi(n) \cdot_2 \pi(m).$$

x5.5. Suppose $(N_1, 0_1, S_1)$ and $(N_2, 0_2, S_2)$ are systems of natural numbers, $\leq_1, \leq_2$ are the respective wellorderings and $\pi : N_1 \rightarrow N_2$ is the canonical isomorphism. Show that π is order preserving, i.e. for all $n, m \in N_1$,

$$n \leq_1 m \iff \pi(n) \leq_2 \pi(m).$$

x5.6. Every subset B of an interval $[0, n]$ is equinumerous with some $[0, m]$, where $m \leq n$. It follows that if A is finite and $B \subseteq A$, then B is finite and $\#(B) \leq \#(A)$.

Every cardinal number is a set; a **finite cardinal** κ is a cardinal number which is a finite set.

x5.7. Prove that for every finite cardinal number κ ,

$$\kappa =_c [0, \#(\kappa)).$$

x5.8. Show that for all n, m , $[0, m) =_c [n, n+m)$ and infer that the union of two finite sets A, B is finite and such that

$$A \cap B = \emptyset \implies \#(A \cup B) = \#(A) + \#(B).$$

It follows that for any two finite cardinals κ, λ ,

$$\#(\kappa + \lambda) = \#(\kappa) + \#(\lambda).$$

x5.9. If $\mathcal{E}$ is a finite set and every member of it is a finite set, then the unionset $\bigcup \mathcal{E}$ is also finite.

x5.10. The product of two finite sets A, B is finite and such that

$$\#(A \times B) = \#(A) \cdot \#(B).$$

It follows that for any two finite cardinals κ, λ ,

$$\#(\kappa \cdot \lambda) = \#(\kappa) \cdot \#(\lambda).$$

x5.11. The powerset of every finite set A is finite and

$$\#(\mathcal{P}(A)) = 2^{\#(A)}.$$

It follows that for every finite cardinal κ ,

$$\#(2^\kappa) = 2^{\#(\kappa)}.$$

x5.12. For all finite cardinals κ, λ ,

$$\#(\kappa^\lambda) = \#(\kappa)^{\#(\lambda)}.$$

x5.13. For all cardinals κ , $2^\kappa \neq \aleph_0$.

x5.14. $\mathfrak{c} + \mathfrak{c} =_c \aleph_0 \cdot \mathfrak{c} =_c \mathfrak{c} \cdot \mathfrak{c} =_c \mathfrak{c}$.

x5.15. $\mathfrak{c}^\mathfrak{c} =_c 2^\mathfrak{c}$.

x5.16. For every cardinal number $\kappa > 1$, if $\kappa \cdot \kappa =_c \kappa$, then $2^\kappa =_c \kappa^\kappa$.

x5.17. For each cardinal number κ and each $n \in \mathbb{N}$,

$$\kappa^n =_c \kappa^{|[0,n]|},$$

where the left side is defined by **5.30** and the right side is cardinal exponentiation.

x5.18. For each $n \neq 0$, $\mathfrak{c}^n =_c |\mathfrak{c}^*| =_c \mathfrak{c}$.

x5.19. For every

$$g : Y \rightarrow E, \quad h : E \times \mathbb{N} \times Y \rightarrow E,$$

there exists exactly one function $f : \mathbb{N} \times Y \rightarrow E$ which satisfies the identities

$$\begin{aligned} f(0, y) &= g(y) & (y \in Y), \\ f(Sn, y) &= h(f(n, y), n, y) & (y \in Y, n \in \mathbb{N}). \end{aligned}$$

x5.20. The function ρ in the proof of **5.31** is a bijection.

***x5.21.** Every partial ordering $\leq$ on a finite set P has a **linearization**, i.e. some linear ordering $\leq'$ of P exists such that $x \leq y \implies x \leq' y$.

***x5.22. The marriage problem.** Suppose B is a finite set and $h : B \rightarrow \mathcal{P}(G)$ is a function, such that for each $x \in B$, $h(x)$ is a finite subset of G and

$$X \subseteq B \implies |X| \leq |\bigcup \{h(x) \mid x \in X\}|, \quad (5.24)$$

so in particular each $h(x) \neq \emptyset$. Prove that there exists a function $f : B \rightarrow G$ such that

$$(\forall x \in B)[f(x) \in h(x)]. \quad (5.25)$$

Show also that both (5.24) and the hypothesis that each $h(x)$ is finite are necessary for the existence of some f which satisfies (5.25). **HINT:** Consider whether or not there exists some $\emptyset \neq C \subseteq B$ such that $|C| = |\bigcup \{h(x) \mid x \in C\}|$. The name of the problem comes from the traditional interpretation that B is a set of boys, G is a set of available girls and h assigns to each boy x the (finite) set $h(x)$ of girls that he would be willing to marry. There are many other applications of the problem, more useful and less sexist, for example when B is a set of courses, G is a set of professors and h assigns to each course the set of professors who can teach it (“the scheduling problem”).

The next problem justifies another form of recursive definition which is often useful in applications.

x5.23. Complete Recursion. For each $h : E^* \rightarrow E$, there exists exactly one function $f : N \rightarrow E$ which satisfies the identity

$$f(n) = h(\bar{f}(n)).$$

The next problem gives a characterization of countable, infinite sets directly in terms of the membership relation, with no appeal to the defined notions of N and “function.”

x5.24. Prove the equivalence:

$$\begin{aligned} A =_c N \iff & (\exists \mathcal{E})[A = \bigcup \mathcal{E} \\ & \& \emptyset \in \mathcal{E} \\ & \& (\forall u \in \mathcal{E})(\exists! y \notin u)[u \cup \{y\} \in \mathcal{E}] \\ & \& (\forall Z)[[\emptyset \in Z \& (\forall u \in Z)(\exists! y \notin u)u \cup \{y\} \in Z \cap \mathcal{E}] \\ & \implies \mathcal{E} \subseteq Z]]. \end{aligned}$$

Chapter 6

FIXED POINTS

The primary significance of the Recursion Theorem 5.6 is foundational, since the result justifies on the basis of the axioms a method of definition of functions which is intuitively obvious. From a purely mathematical point of view, however, we can also view 5.6 as a theorem of *existence and uniqueness* of solutions for systems of identities of the form

$$\begin{aligned} f(0) &= a, \\ f(Sx) &= h(f(x)) \quad (x \in N), \end{aligned} \tag{6.1}$$

where $a \in E$ and $h : E \rightarrow E$ are given and the function $f : N \rightarrow E$ is *the unknown*. In this chapter we will prove an elegant generalization of the Recursion Theorem in the context of the theory of partial orderings, which implies the existence and uniqueness of solutions for systems of functional identities much more general than (6.1). The CONTINUOUS LEAST FIXED POINT THEOREM 6.21 is fundamental for the *theory of computation*, it is the basic mathematical tool of the so-called *fixpoint theory of programs*. In the next chapter we will show that it is a special case of a much deeper FIXED POINT THEOREM of Zermelo, which is intimately related to the theory of wellorderings and rich in set theoretic consequences, for example it implies directly the Hypothesis of Cardinal Comparability, 3.1. Thus, in addition to its purely mathematical significance, the Continuous Least Fixed Point Theorem yields also an interesting point of contact between classical set theory and today's theoretical computer science.

In their simplest and most natural expressions, the Fixed Point Theorems are somewhat abstract and apparently unrelated to the solution of systems of functional identities to which we intend to apply them. To understand what they say and how to use them, we will need to introduce first some basic notions from the theory of partial orderings.

6.1. A partially ordered set or simply **poset** is a structured set

$$P = (Field(P), \leq_P),$$

where $Field(P)$ is an arbitrary set and $\leq_P$ is a partial ordering on $Field(P)$, i.e. a reflexive, transitive and antisymmetric binary relation. Notice that



Figure 6.1. A discrete and a flat poset.

$\leq_P$ determines P because it is reflexive,

$$x \in \text{Field}(P) \iff x \leq_P x,$$

so we can specify a poset P completely by defining its partial ordering $\leq_P$. In practice, however, the partial ordering $\leq_P$ is often clear from the context and we will tend to identify a poset P with its field $\text{Field}(P)$, following the general convention about structured sets discussed in 4.26. For example, by *the poset* N we obviously mean the pair $(N, \leq)$, where $\leq$ is the usual ordering on the set of natural numbers. By this convention, *the points of* P are the members of $\text{Field}(P)$, a *subset* $I \subseteq P$ is a subset $I \subseteq \text{Field}(P)$, etc. Each $I \subseteq P$ is a poset in its own right, partially ordered by the restriction of $\leq_P$ to I ,

$$x \leq_I y \iff_{\text{df}} x \leq_P y \ \& \ x \in I \ \& \ y \in I, \quad (6.2)$$

which is (easily) a partial ordering. We will often deal with posets which have a least element, and it will be convenient to use the same, standard symbol $\perp$ (read “bottom” or “least”) for it, just as we use the same symbol 0 for the additive unit of every number system:

$$\perp = \perp_P =_{\text{df}} \text{the least element of } P \text{ (if it exists)}. \quad (6.3)$$

Any set A can be viewed as a **discrete poset** in which no two elements are comparable, i.e. partially ordered by the identity relation

$$x \leq y \iff x = y \quad (x, y \in A).$$

Just above these in complexity are the **flat posets** which have a least element, the only element involved in any comparisons: i.e.

$$x \leq_P y \iff x = \perp \vee x = y.$$

The simplest non-empty poset is a singleton $\{\perp\}$, which is both discrete and flat. Additional examples of posets are the set N of natural numbers (with its usual ordering), as well as the sets Q and R of the rationals and the reals which we have not yet defined carefully within axiomatic set theory. These are all linear (totally ordered) posets. There is a large variety of finite posets and their study constitutes an important research area of mathematics, but we will not be much concerned with it here. Mostly we will use them as counterexamples. In drawing posets we indicate $x < y$ by

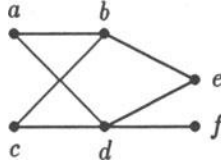


Figure 6.2. A finite poset

placing y above or to the right of x and drawing a line from x to y , which may pass through other points, e.g. $c < e$ in Figure 6.2.¹

6.2. Definition. Let P be a poset, $S \subseteq P$ and $M \in P$ a member of P .

1. M is an **upper bound** of S if it is greater than or equal to every element of S , $(\forall x \in S)[x \leq M]$.
2. M is **maximum** in S if it is a member and an upper bound of S , i.e. $M \in S$ & $(\forall x \in S)[x \leq M]$.
3. M is a **least upper bound** of S if it is an upper bound and also less than or equal to every other upper bound of S , i.e.

$$(\forall x \in S)[x \leq M] \text{ \& \& } (\forall M')[(\forall x \in S)[x \leq M'] \implies M \leq M'].$$

If M_1, M_2 are both least upper bounds of S , then $M_1 \leq M_2$ (because M_2 is an upper bound and M_1 is a least upper bound) and symmetrically $M_2 \leq M_1$, so that $M_1 = M_2$, i.e. S has at most one least upper bound. When it exists, the least upper bound of a set S is denoted by

$$\sup S = \text{the least upper bound of } S. \quad (6.4)$$

The term “*sup*” from the Latin *supremum* (maximum) is justified by the following observation.

6.3. Exercise. If M is maximum of a set S in a poset P , then M is also the least upper bound of S .

6.4. Exercise. In the poset of Figure 6.2, find subsets S with the following properties: (1) S has no upper bound. (2) S has upper bounds but no least upper bound. (3) S has a least upper bound but no maximum element.

¹There are those who draw posets growing to the right, those who draw them growing up and even those who draw them growing down; to the best of my knowledge, nobody pictures posets growing to the left and it does not appear that any of the three common choices is dominant.

6.5. Exercise. In any poset P , an element M is the least upper bound of the empty set $\emptyset$ if and only if M is the least element of P , i.e.

$$\perp = \sup \emptyset \quad (6.5)$$

if $\perp$ or $\sup \emptyset$ exists.

6.6. Exercise. The powerset $\mathcal{P}(A)$ of every set A is partially ordered by the relation

$$X \subseteq Y \iff_{\text{df}} X \subseteq Y \subseteq A,$$

so that $\perp = \emptyset$ and for every $S \subseteq \mathcal{P}(A)$, the union $\bigcup S$ is the least upper bound of S .²

Less trivial and more interesting for our purposes is the next example of a poset.

6.7. Definition. A **partial function** on a set A to a set E is any function with domain of definition some subset of A and values in E , in symbols

$$f : A \rightarrow E \iff_{\text{df}} \text{Function}(f) \ \& \ \text{Domain}(f) \subseteq A \ \& \ \text{Image}(f) \subseteq E. \quad (6.6)$$

For example, $(n \mapsto (n - 1))$ is a partial function on the natural numbers defined only when $n \neq 0$, $(x \mapsto \sqrt{x})$ is a partial function on the reals with domain of definition $\{x \mid x \geq 0\}$, etc. A finite sequence $u \in A^*$ is a partial function $u : N \rightarrow A$, as we defined it in 5.29. The empty set $\emptyset$ is (trivially) a partial function (with empty domain of definition!) and every (total) function on A to E is also a partial function, since (6.6) does not exclude $\text{Domain}(f) = A$,

$$\emptyset : A \rightarrow E, \quad f : A \rightarrow E \implies f : A \rightarrow E.$$

We will use systematically the convenient half-arrow notation for partial functions (recently established in computer science), as well as the common notations

$$f(x) \downarrow \iff_{\text{df}} x \in \text{Domain}(f), \quad f(x) \uparrow \iff_{\text{df}} x \notin \text{Domain}(f) \quad (6.7)$$

for indicating that a partial function is defined or undefined at some point.

6.8. Definition. For each A and E ,

$$(A \rightarrow E) =_{\text{df}} \{f \subseteq A \times E \mid f : A \rightarrow E\} \quad (6.8)$$

²More pedantically, the partial ordering of $\mathcal{P}(A)$ is the restriction $\subseteq_A$ of the definite condition $X \subseteq Y$ to $\mathcal{P}(A)$, (4.11).

is the set of all partial functions from A to E , in analogy with the notation $(A \rightarrow E)$ for the set of all (total) functions from A to E , (4.22). The set $(A \multimap E)$ is partially ordered by the relation $\subseteq$,

$$f \subseteq g \iff (\forall x \in A)[f(x) \downarrow \implies [g(x) \downarrow \ \& \ f(x) = g(x)]],$$

with least element $\perp = \emptyset$.

6.9. Exercise. For each A, E ,

$$(A \multimap E) = \{f \upharpoonright X \mid f : X \rightarrow E \ \& \ X \subseteq A\}.$$

Function restrictions are defined in (4.21).

It is harder to find least upper bounds in these partial function posets than in powersets: for example, the set of two constant functions $\{x \mapsto 0, x \mapsto 1\}$ in $(N \multimap N)$ has no upper bound at all, because any partial function above both $(x \mapsto 0)$ and $(x \mapsto 1)$ would need to satisfy the contradictory $h(0) = 0$ and $h(0) = 1$. On the other hand, linearly ordered subsets of $(A \multimap E)$ have least upper bounds and this is a fruitful property of these posets, worth a name.

6.10. Definition. A **chain** in a poset P is any linearly ordered subset S of P , i.e. a subset satisfying

$$(\forall x, y \in S)[x \leq y \vee y \leq x].$$

A poset P is **chain-complete** or **inductive** if every chain in P has a least upper bound.

6.11. Exercise. The empty set is (trivially) a chain, hence every inductive poset P has a least element $\perp = \sup \emptyset$.

6.12. Exercise. Every flat poset is inductive; a discrete poset is inductive only when it has exactly one element, in which case it is also flat.

6.13. Exercise. The image $\{x_n \mid n \in N\}$ of a non-decreasing sequence

$$x_0 \leq x_1 \leq x_2 \leq \dots$$

is a chain; thus, every non-decreasing sequence has a **limit** in an inductive poset,

$$\lim_n x_n =_{\text{df}} \sup \{x_n \mid n \in N\}. \quad (6.9)$$

6.14. Proposition. (1) For each set A , the powerset $\mathcal{P}(A)$ is inductive. (2) For any two sets A, E , the poset $(A \rightarrow E)$ of all partial functions from A to E is inductive. (3) For every poset P , the set

$$\text{Chains}(P) = \{S \subseteq P \mid S \text{ is a chain}\}$$

of all chains in P (partially ordered under $\subseteq$) is inductive.

Proof. (2) If $S \subseteq (A \rightarrow E)$ is a chain, then the union $\bigcup S$ is a partial function and obviously, $\bigcup S = \sup S$. (3) This is also proved by observing that the union of a chain of chains in a poset is also a chain. $\dashv$

6.15. Exercise. Neither N nor the finite poset of Figure 6.2 are inductive.

6.16. Exercise. For each set E , the set $P = E^* \cup (N \rightarrow E)$ of finite and infinite sequences from E is an inductive poset, under $\subseteq$.

6.17. Exercise. For any two sets A, E , the poset

$$(A \hookrightarrow E) =_{\text{df}} \{f \in (A \rightarrow E) \mid f \text{ is one-to-one}\}$$

of partial injections from A to E (partially ordered by $\subseteq$) is inductive.

We will find the most significant applications of the fixed point theorems in partial function posets, but the proofs will use only the fact that they are inductive, and there are lots of other interesting examples. Some of them are described in the problems at the end of the chapter.

Finally, we need to delineate the type of functions on inductive posets which must, necessarily, have fixed points.

6.18. Definition. A mapping³ $\pi : P \rightarrow Q$ on a poset P to another is **monotone** if for all $x, y \in P$,

$$x \leq_P y \implies \pi(x) \leq_Q \pi(y).$$

A monotone mapping need not be strictly increasing in the sense of

$$x <_P y \implies \pi(x) <_Q \pi(y),$$

e.g. every constant mapping is monotone.

³It is convenient to refer to $\pi : P \rightarrow Q$ as a “mapping” rather than a “function” (which means the same thing), because in the interesting applications P is some partial function space $(A \rightarrow E)$, Q may be another partial function space, π takes partial functions as arguments and (possibly) values and there are altogether too many functions around. Notice also that, pedantically, $\pi : \text{Field}(P) \rightarrow \text{Field}(Q)$ is a mapping from the field of P to that of Q .

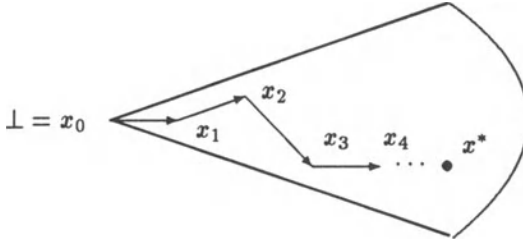


Figure 6.3. The Continuous Least Fixed Point Theorem.

Notice that if $\pi : P \rightarrow Q$ is monotone and $S \subseteq P$ is a chain, then the image $\pi[S]$ is also a chain; because given $x = \pi(u)$, $y = \pi(v)$ with $u, v \in S$, either $u \leq v$, which implies $x = \pi(u) \leq \pi(v) = y$, or $v \leq u$, which similarly implies $y \leq x$. This makes the next definition meaningful.

6.19. Definition. A monotone mapping $\pi : P \rightarrow Q$ on an inductive poset to another is **countably continuous** if for every non-empty, countable chain $S \subseteq P$,

$$\pi(\sup S) = \sup \pi[S].$$

6.20. Exercise. A monotone mapping $\pi : P \rightarrow Q$ on one inductive poset to another is countably continuous if and only if for every non-decreasing sequence $x_0 \leq_P x_1 \leq_P \dots$ of elements in P ,

$$\pi(\lim_n x_n) = \lim_n \pi(x_n).$$

Here the limit on the left is taken in P and the limit on the right is taken in Q .

6.21. Continuous Least Fixed Point Theorem. Every countably continuous, monotone mapping $\pi : P \rightarrow P$ on an inductive poset into itself has exactly one **strongly least fixed point** x^* , which is characterized by the two properties

$$\pi(x^*) = x^*, \tag{6.10}$$

$$(\forall y \in P)[\pi(y) \leq y \implies x^* \leq y]. \tag{6.11}$$

Proof. The **orbit** of the least element $\perp$ under π is defined by the simple recursion on the natural numbers,

$$\begin{aligned} x_0 &= \perp, \\ x_{n+1} &= \pi(x_n). \end{aligned}$$

Clearly $x_0 = \perp \leq x_1$, and by a trivial induction (using the monotonicity of π), for every n , $x_n \leq x_{n+1}$. Thus, the limit

$$x^* =_{\text{df}} \lim_n x_n = \sup \{x_n \mid n \in N\} \quad (6.12)$$

exists by **6.13**, and by the countable continuity of π ,

$$\pi(x^*) = \pi(\lim_n x_n) = \lim_n \pi(x_n) = \lim_n x_{n+1} = x^*.$$

For the second claimed property of x^* , we assume $\pi(y) \leq y$ and show by induction that for every n , $x_n \leq y$. **BASIS.** $x_0 = \perp \leq y$, trivially. **INDUCTION STEP.** The Induction Hypothesis gives us $x_n \leq y$, and we compute:

$$\begin{aligned} x_n \leq y &\implies \pi(x_n) \leq \pi(y), && \text{because } \pi \text{ is monotone,} \\ &\implies x_{n+1} \leq \pi(y) \leq y, && \text{by the assumption on } y. \end{aligned}$$

Thus, y is an upper bound of the chain $\{x_n \mid n \in N\}$, and hence, $x^* = \sup \{x_n \mid n \in N\} \leq y$. $\dashv$

To apply the Continuous Least Fixed Point Theorem, we must formulate the problem at hand as a question of existence and (sometimes) uniqueness of solutions for an equation of the form $\pi(x) = x$, where $\pi : P \rightarrow P$ is monotone and countably continuous on some inductive poset P . This is typically the hardest part: to bring the problem in a form in which **6.21** can be applied. *Verification of the countable continuity of π is not necessary:* because we will show in the next chapter that **6.21** remains true if we simply remove the hypothesis of countable continuity of π . In any case, most applications involve simple monotone mappings on partial function posets for which it is often trivial to recognize a much stronger, natural continuity property.

6.22. Definition. A partial function $g : A \rightarrow E$ is **finite** if it has finite domain, i.e. if it is a finite set of ordered pairs. A mapping $\pi : (A \rightarrow E) \rightarrow (B \rightarrow M)$ from one partial function space into another is **continuous**, if it is monotone and for each $f : A \rightarrow E$, and each $y \in B$ and $v \in M$,

$$\pi(f)(y) = v \implies (\exists g \subseteq f)[g \text{ is finite \& } \pi(g)(y) = v]. \quad (6.13)$$

The notation is a bit convoluted but what it means is quite simple: to compute $\pi(f)(y)$, we first compute the partial function $f' = \pi(f)$ and then we evaluate it at y , $\pi(f)(y) = f'(y)$; if π is continuous, then each value $\pi(f)(y)$ of $\pi(f)$, whenever defined, depends only on finitely many values of f . The continuity of specific mappings is often obvious, by inspection of their definition: we simply need to notice that each value $\pi(f)(y)$ (when

defined) is determined from a finite number of values of f . For example, the mapping $\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$ defined by

$$\pi(f) = (n \mapsto f(n) + f(n^2))$$

is obviously continuous, since each $\pi(f)(n) = f(n) + f(n^2)$ depends only on two values of f .

6.23. Exercise. Show that the mapping $\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$ defined by

$$\pi(f) = (n \mapsto \sum_{i=0}^n f(i))$$

is continuous. Compute $\pi(n \mapsto 2n)(2)$ for this π .

6.24. Definition. A function $f : X \rightarrow Y$ from one topological space to another is (topologically) **continuous**, if the inverse image $f^{-1}[G]$ of every open subset of Y is an open subset of X . The definition of a **topological space** was sneaked in 4.26, as the first example of a structured set.

6.25. Exercise. A function $f : X \rightarrow Y$ from one topological space to another is continuous if and only if the inverse image $f^{-1}[F]$ of every closed subset of Y is closed in X .

One might guess that our using the term “continuous” in Definition 6.22 is not entirely accidental and that the notion of 6.22 has something to do with topological continuity. Indeed it does, the notions are equivalent when the proper topology is put on partial function posets, but we will have no need of this fact and will leave it for Problem x6.19.

6.26. About topology. General (pointset) topology is to set theory like parsley to Greek food, some of it gets in almost every dish, but there are no great “parsley recipes” that the good Greek cook needs to know. Many notions and results of set theory are connected to topological ideas, but it is quite rare that you can prove an interesting theorem about sets by quoting some deep topological result. To avoid getting distracted with side issues, we will follow the general policy of giving the most direct, set theoretically natural definitions and proofs of the notions and results we need and leave the connections with topology for the problems. Occasionally the most natural approach is topological.

6.27. Lemma. If $S \subseteq (A \rightarrow E)$ is a non-empty chain in a partial function poset and $g \subseteq \sup S$ is a finite function, then there exists some $h \in S$ such that $g \subseteq h$.

Proof is by induction on the number of elements in the domain of g . **BASIS.** $g = \emptyset$ is the partial function which is nowhere defined. There is some $h \in S$

since S is non-empty, and $\emptyset \subseteq h$. INDUCTION STEP. The domain of g has $n + 1$ elements, so

$$g = g_1 \cup \{(x, w)\} \subseteq \sup S,$$

where g_1 is a finite, partial function with just n elements in its domain, and by induction hypothesis, there exists some $h_1 \in S$ such that $g_1 \subseteq h_1$. Since $(x, w) \in \sup S$, there must also exist some $h' \in S$ such that $(x, w) \in h'$, and since S is a chain, either $h_1 \subseteq h'$ or $h' \subseteq h_1$; the h we need is the larger of these two partial functions. $\dashv$

6.28. Lemma. *Every continuous mapping $\pi : (A \rightarrow E) \rightarrow (B \rightarrow M)$ is countably continuous, in fact, for every (not necessarily countable) non-empty chain $S \subseteq (A \rightarrow E)$,*

$$\pi(\sup S) = \sup \pi[S].$$

Proof. We must show that if $S \subseteq (A \rightarrow E)$ is a non-empty chain with union $f = \sup S$ and $\pi(f)(y) = v$, then there exists some $h \in S$ such that $\pi(h)(y) = v$. By the strong continuity of π , there exists a finite $g \subseteq f$, such that already $\pi(g)(y) = v$; by the preceding Lemma, there exists some $h \in S$ so that $g \subseteq h$; and by the monotonicity of π , this implies $\pi(g) \subseteq \pi(h)$. In particular, since $\pi(g)(y) = v$, we have $\pi(h)(y) = v$, so this is the h we need. $\dashv$

The Continuous Least Fixed Point Theorem is evidently a simple corollary of the Recursion Theorem on the natural numbers 5.6. In fact, it implies 5.6 by a fairly direct argument, which is worth looking at, as it illustrates how we intend to apply 6.21.

6.29. Proof of the Recursion Theorem from 6.21. For each given $a \in E$ and function $h : E \rightarrow E$, we define the mapping

$$\pi : (N \rightarrow E) \rightarrow (N \rightarrow E)$$

by the formula

$$\pi(f) = f', \text{ where } f'(x) = \begin{cases} a, & \text{if } x = 0, \\ h(f(x-1)), & \text{if } x > 0, \end{cases}$$

where f is any partial function from N to E and we understand the definition naturally, so that

$$x > 0 \implies [f'(x) \downarrow \iff h(f(x-1)) \downarrow \iff f(x-1) \downarrow].$$

Written out in detail, the mapping π associates a set of pairs $f' \subseteq (N \times E)$ with every $f \in (N \rightarrow E)$ and it is defined by the equation

$$\begin{aligned} \pi(f) = \{ (0, a) \} \\ \cup \{ (x, h(w)) \mid x > 0 \ \& \ (x-1, w) \in f \} \quad (f : N \rightarrow E). \end{aligned} \tag{6.14}$$

This formal definition makes it quite obvious that π is monotone—just imagine substituting for f in (6.14) a larger $f' \supseteq f$ and see that you get a larger $\pi(f')$. It also makes it clear that π is continuous, since each pair $(x, h(w)) \in \pi(f)$ (other than $(0, a)$ which we throw in for free) is included because the single point $(x-1, w) \in f$. (In full detail: if $\pi(f)(x) = v$ and $x = 0$, take $g = \emptyset$ in the definition of strong continuity, and if $x > 0$, take $g = \{(x-1, w)\} \subseteq f$, where $v = h(w)$.) It follows that π is countably continuous, so by 6.21 it has a fixed point: that is, some partial function $f^* : N \rightarrow E$ exists which satisfies $f^* = \pi(f^*)$, so that, immediately,

$$f^*(0) = a, \quad (6.15)$$

$$f^*(x+1) = h(f^*(x)) \quad (f^*(x) \downarrow). \quad (6.16)$$

Theorem 6.21 does not guarantee that this f^* is a *total function*, with domain of definition the whole N , but this can be verified by an easy induction on x using the identities (6.15) and (6.16). $\dashv$

Consider next a case where it is not quite so obvious how to define the function we want directly by the Recursion Theorem.

6.30. Proposition. *For each function $h : N \rightarrow N$ and each infinite set $A \subseteq N$ of natural numbers, there exists a (total) function $f : N \rightarrow N$ which satisfies the identity*

$$f(n) = \begin{cases} 0, & \text{if } n \in A, \\ h(f(n+1)), & \text{if } n \notin A. \end{cases} \quad (6.17)$$

Proof. We define the mapping

$$\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$$

on the inductive poset of all unary, partial functions on N by the formula

$$\pi(f) = f', \text{ where } f'(n) = \begin{cases} 0, & \text{if } n \in A, \\ h(f(n+1)), & \text{if } n \notin A. \end{cases} \quad (6.18)$$

In full detail, this means we set

$$\pi(f) = \{(n, 0) \mid n \in A\} \cup \{(n, h(w)) \mid n \notin A \ \& \ (n+1, w) \in f\},$$

which implies by inspection that π is continuous, hence countably continuous. Thus we must have a fixed point f which satisfies (6.17), and it is enough to prove that this f is total. Assume towards a contradiction that $f(n) \uparrow$ for some n . Notice that by (6.17), this means that $n \notin A$, else $f(n) \downarrow$, in fact, $f(n) = 0$. We will prove by induction on i that $f(n+i) \uparrow$, which implies again that for all i , $n+i \notin A$, so that $A \subseteq [0, n)$ is finite,

contradicting the hypothesis. BASIS. If $i = 0$, then $f(n + 0) = f(n) \uparrow$, by assumption. INDUCTION STEP. Assume that $f(n + i) \uparrow$, so that by (6.17), once more, $n + i \notin A$. Now this implies that $f(n + i) = h(f(n + i + 1))$ so that $f(n + i + 1) \downarrow \implies f(n + 1) \downarrow$ (since h is total), which violates the induction hypothesis. $\neg$

6.31. Exercise. Prove in detail that the mapping π in this proof is continuous.

As a third, typical application of the Continuous Least Fixed Point Theorem we consider the *Euclidean algorithm*.

6.32. Proposition. (1) *There exists exactly one partial function $f : N \times N \rightarrow N$ with domain of definition $\{(n, m) \mid n, m \neq 0\}$ which satisfies the following identities for all $0 < n < m$:*

$$\begin{aligned} f(m, n) &= f(n, m), \\ f(n, n) &= n, \\ f(n, m) &= f(n, m - n). \end{aligned} \tag{6.19}$$

(2) *The unique f^* which satisfies the system (6.19) computes the greatest common divisor of any two natural numbers different from 0,*

$$\begin{aligned} f^*(n, m) &= \gcd(n, m) \\ &= \text{the largest } k \text{ which divides evenly} \\ &\quad \text{both natural numbers } n, m. \end{aligned} \tag{6.20}$$

Proof. With each partial function $f : N \times N \rightarrow N$ we associate the partial function $f' : N \times N \rightarrow N$ which is defined by the formula

$$f'(n, m) = \begin{cases} f(m, n), & \text{if } n > m > 0, \\ n, & \text{if } n = m > 0, \\ f(n, m - n) & \text{if } 0 < n < m, \end{cases}$$

and we set

$$\pi(f) = f'.$$

The mapping $\pi : ((N \times N) \rightarrow N) \rightarrow ((N \times N) \rightarrow N)$ is obviously continuous. It follows that there exists a least partial function $f^* : (N \times N) \rightarrow N$ which satisfies

$$\pi(f^*) = f^*,$$

and this is (easily) equivalent with the system (6.19). Proof that for all $n, m \neq 0$

$$f^*(n, m) \downarrow \ \& \ f^*(n, m) = \gcd(n, m)$$

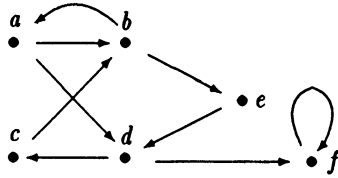


Figure 6.4.

is by induction on the sum $n + m$. (Take cases whether $n > m > 0$, $n = m > 0$ or $0 < n < m$, and use the simple property of the natural numbers, that for $0 < n < m$, the common divisors of n, m are precisely the same as the common divisors of $n, m - n$.) $\dashv$

In this example we do not need the Least Fixed Point Theorem to prove the existence of a solution for the system (6.19), since we can verify directly that the function $\gcd$ is a solution. Despite this, the proposition is important because it yields a characterization of the function $\gcd$ which suggests a specific—and simple—method for computing it. For example, using only the identities of the system, we compute:

$$\begin{aligned} \gcd(231, 165) &= \gcd(165, 231) = \gcd(165, 66) \\ &= \gcd(66, 165) = \gcd(66, 99) = \gcd(66, 33) \\ &= \gcd(33, 66) = \gcd(33, 33) = 33. \end{aligned}$$

This computation of the value $\gcd(231, 165)$ is much simpler than the trivial one, where we would search for the greatest common divisor by testing in sequence all the numbers from 165 moving down, until we would find some common divisor of 165 and 231. The example is quite general: the characterization of a partial function f as the least solution of a system of simple identities typically yields an **algorithm**, a “recipe” for the “mechanical” computation of the values of f , and this is the underlying reason for the significance of the Continuous Least Fixed Point Theorem in theoretical computer science.

We end with a simple result about graphs which is related to the ideas of this chapter, see Problems **x6.15** and **x6.16**.

6.33. Definition. A **graph** is a structured set $(G, \rightarrow_G)$, where the set of **edges** $\rightarrow_G \subseteq G \times G$ is an arbitrary binary relation on the set of **nodes** G . The **transitive closure** of a graph G is the graph $\overline{G} = (G, \Rightarrow_G)$, where

$$\begin{aligned} x \Rightarrow_G y &\iff_{\text{df}} \text{there is a path from } x \text{ to } y \text{ in } G \\ &\iff (\exists z_0, \dots, z_n)[x = z_0 \rightarrow_G z_1 \ \& \ \dots \ \& \ z_{n-1} \rightarrow_G z_n = y]. \end{aligned}$$

We draw graphs much like posets, except we forget about the convention of “growing up or towards the right” and use arrows instead of lines: $x \rightarrow_G y$

holds if there is an arrow from x to y , and $x \Rightarrow_G y$ holds if you can move from x to y along the arrows of the diagram. In Figure 6.4 we have $f \rightarrow f$, $a \Rightarrow a$ and $a \Rightarrow c$, but $f \not\Rightarrow d$.

6.34. Proposition. *For each graph G , the transitive closure relation $\Rightarrow_G$ satisfies the equivalence*

$$x \Rightarrow_G y \iff x \rightarrow_G y \vee (\exists z \in G)[x \rightarrow_G z \ \& \ z \Rightarrow_G y]. \quad (6.21)$$

Proof. Suppose first that (skipping the conjunction signs)

$$x = z_0 \rightarrow_G z_1 \rightarrow_G z_2 \rightarrow_G \cdots \rightarrow_G z_n = y;$$

if $n = 1$, we have $x \rightarrow_G y$, and if $n > 1$, then $x \rightarrow_G z_1$ and $z_1 \Rightarrow_G y$ (by the definition of $\Rightarrow_G$), so we have the right-hand side of (6.21), taking $z = z_1$. The converse is equally simple, taking cases on the two disjuncts of the right-hand side. $\dashv$

Problems

x6.1. For every partial ordering $\leq$ on a set A , the **converse relation**

$$x \leq' y \iff_{\text{df}} y \leq x$$

is also a partial ordering. Of the inductive posets $(A \rightarrow E)$ and $\mathcal{P}(A)$, which one has an inductive, converse poset?

x6.2. Suppose $\leq_E$ is an inductive partial ordering on the set E , A is a set and $\leq$ is the “pointwise” partial ordering on the function space $(A \rightarrow E)$,

$$f \leq g \iff_{\text{df}} (\forall x \in A)[f(x) \leq_E g(x)] \quad (f, g : A \rightarrow E).$$

Prove that $\leq$ is an inductive partial ordering on $(A \rightarrow E)$.

x6.3. If the partial orderings $\leq_1, \leq_2$ on the respective sets P_1, P_2 are inductive, then the following relation $\leq$ on the Cartesian product $P_1 \times P_2$ is also inductive:

$$(x_1, x_2) \leq (y_1, y_2) \iff_{\text{df}} x_1 \leq_1 y_1 \ \& \ x_2 \leq_2 y_2.$$

With this partial ordering, the poset $P_1 \times P_2$ is called **the product** of the two posets P_1 and P_2 .

x6.4. Suppose P_1, P_2, Q are inductive posets. A mapping $\pi : P_1 \times P_2 \rightarrow Q$ is **separately monotone** if for each $x_1 \in P_1$, the mapping $(x_2 \mapsto \pi(x_1, x_2))$ is monotone, and symmetrically for each $x_2 \in P_2$; π is **separately, countably continuous** if for each $x_1 \in P_1$, the mapping $(x_2 \mapsto \pi(x_1, x_2))$ is countably continuous, and symmetrically for each $x_2 \in P_2$. Prove that π is monotone (on the product poset) if and only if it is separately monotone, and countably continuous if and only if it is separately countably continuous.

6.35. Definition. A point M is **maximal** in a subset S of a poset P if M is a member of S and no member of S is bigger,

$$M \in S \ \& \ (\forall x \in S)[M \leq x \implies M = x].$$

A point m is **minimal** in S if it is a member of S and no member of S is smaller,

$$m \in S \ \& \ (\forall x \in S)[x \leq m \implies x = m].$$

x6.5. Find in the poset of Figure 6.2 a subset S which has a maximal element but no maximum and another subset S' which has a minimal element but no minimum.

***x6.6.** Every finite, non-empty subset of an arbitrary poset P has at least one maximal and one minimal member.

***x6.7.** A finite poset P is inductive if and only if it has a least element.

An important notion in computer science is that of a *stream*, for example the stream of bytes in a file transmitted over the telephone lines to my home computer from the University of Athens CYBER. A stream is basically a sequence, but it may be *infinite*, in the idealized case; *terminated*, if after some stage an end-of-file signal comes and my machine knows that the transmission is done; or *unterminated*, if after some stage the bytes stop coming, without warning, perhaps because the CYBER died or the telephone connection was interrupted.

6.36. Definition. For each set A , we fix some $t \notin A$ (for example, the object $r(A)$ of (3.4)) and we define the **streams from A** by:

$$\text{Streams}(A) =_{\text{df}} \{ \sigma : N \rightarrow A \cup \{t\} \mid (\forall i < j)[\sigma(j) \downarrow \implies [\sigma(i) \downarrow \ \& \ \sigma(i) \neq t]] \}.$$

We call a stream σ **terminated** or **convergent** if for some n , $\sigma(n) = t$, in which case, by the definition $\text{Domain}(\sigma) = [0, n + 1)$; **infinite** if $\text{Domain}(\sigma) = N$; and **unterminated** if $\text{Domain}(\sigma)$ is a finite, initial segment of N but σ does not take on the *terminating value* t . The infinite and unterminated streams together are called **divergent**.

x6.8. For each set A , the set of streams $Streams(A)$ is an inductive poset under the natural, partial ordering $\sqsubseteq$, where, as for strings,

$$\sigma \sqsubseteq \tau \iff \sigma \subseteq \tau. \quad (6.22)$$

What are its maximal elements?

x6.9. The **concatenation** of two streams is defined so that if σ is divergent, then $\sigma \star \tau = \sigma$ and if σ is convergent with domain $[0, n + 1)$, then

$$i < n \implies \sigma \star \tau(i) = \sigma(i), \quad \sigma \star \tau(n + i) = \tau(i).$$

Prove that $\star$ is a continuous function (of two variables) on $Streams(A)$.

The full Least Fixed Point Theorem can be proved directly and easily for powersets:

***x6.10.** Suppose $\pi : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ is a monotone mapping on a powerset. Prove that the set

$$A^* = \bigcap \{X \mid \pi(X) \subseteq X\}$$

is the least fixed point of π , and

$$A_* = \bigcup \{X \mid X \subseteq \pi(X)\}$$

is the largest fixed point of π .

The next few problems deal with “algorithmic” applications of the Least Fixed Point Theorem.

x6.11. For each relation $R \subseteq N \times A$, there exists a least partial function $f : N \times A \rightarrow N$ such that

$$\begin{cases} nRx & \implies f(n, x) = n, \\ \neg nRx & \implies f(n, x) = f(n + 1, x), \end{cases}$$

It follows that

$$\begin{aligned} f(n, x) \downarrow & \iff (\exists m \geq n)[mRx], \\ f(n, x) \downarrow & \implies f(n, x) = \text{the least } m \geq n \text{ such that } [mRx]. \end{aligned}$$

x6.12. For any three partial functions f_0, g, h with domains and ranges such that the identities below make sense, there exists a least partial function $f : N \times A \rightarrow E$ which satisfies the identities

$$\begin{aligned} f(0, x) &= f_0(x), \\ f(n + 1, x) &= h(f(n, g(n, x)), n, x). \end{aligned}$$

x6.13. Prove that there is exactly one total function $f : N \times N \rightarrow N$ which satisfies the identities

$$\begin{aligned} f(0, n) = f(n, 0) &= 0, \\ f(n+1, m+1) &= f(n, m) + 1. \end{aligned}$$

Compute $f(5, 23)$ using these identities and “explain” what $f(n, m)$ is, for any n, m .

6.37. Definition. On the set E^* of strings (finite sequences) from a set E defined in 5.29, we define the partial functions

$$\text{head}(u) = u(0), \quad (6.23)$$

$$\text{tail}(u) = \langle u(1), \dots, u(\text{lh}(u) - 1) \rangle. \quad (6.24)$$

Notice that $\text{head}(u) \downarrow$ when $\text{lh}(u) > 0$, and $\text{tail}(u)$ is always defined, but it is the empty string when $\text{lh}(u) \leq 1$.

x6.14. Prove that there exists a unique, total function $r : E^* \rightarrow E^*$ which satisfies the identity

$$r(u) = \begin{cases} u, & \text{if } \text{lh}(u) \leq 1, \\ r(\text{tail}(u)) \star \langle \text{head}(u) \rangle, & \text{if } \text{lh}(u) > 1. \end{cases}$$

Compute $r(\langle a, b, c \rangle)$ and describe $r(u)$ in general.

x6.15. Prove that for each graph G , the relation $\Rightarrow_G$ is the least (under $\subseteq$) transitive relation on G which includes the edge relation $\rightarrow_G$.

x6.16. Prove that for every graph G with edge relation $\rightarrow_G$, the relation $\Rightarrow_G$ is the common least fixed point of the following monotone operators on the poset $\mathcal{P}(G \times G)$ of all binary relations on G :

$$\begin{aligned} \pi_1(R) &= \{(x, y) \mid x \rightarrow_G y \vee (\exists z)[x \rightarrow_G z \ \& \ (z, y) \in R]\}, \\ \pi_2(R) &= \{(x, y) \mid x \rightarrow_G y \vee (\exists z)[(x, z) \in R \ \& \ z \rightarrow_G y]\}, \\ \pi_3(R) &= \{(x, y) \mid x \rightarrow_G y \vee (\exists z)[x \rightarrow_G z \rightarrow_G y] \\ &\quad \vee (\exists z, w)[x \rightarrow_G z \ \& \ (z, w) \in R \ \& \ w \rightarrow_G y]\}. \end{aligned}$$

x6.17. Let P_1, P_2 be inductive posets and

$$\begin{aligned} \pi_1 : P_1 \times P_2 &\rightarrow P_1, \\ \pi_2 : P_1 \times P_2 &\rightarrow P_2 \end{aligned}$$

arbitrary countably continuous, monotone mappings, where $P_1 \times P_2$ is the product. Prove that there exist unique **least**, **mutual** or **simultaneous fixed points** x_1^*, x_2^* which are characterized by the properties:

$$\pi_1(x_1^*, x_2^*) = x_1^*, \quad \pi_2(x_1^*, x_2^*) = x_2^*,$$

$$\pi_1(y_1, y_2) \leq_1 y_1 \ \& \ \pi_2(y_1, y_2) \leq_2 y_2 \implies x_1^* \leq_1 y_1 \ \& \ x_2^* \leq_2 y_2.$$

The next problem is an algorithmic version of the well-known number theoretic result, that for any two natural numbers $n, m \neq 0$, there exist (positive or negative) integers α, β such that

$$\gcd(n, m) = \alpha n + \beta m.$$

The proof uses some simple properties of the set

$$Z = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

of *rational integers* whose faithful representation in set theory we will develop in Appendix A.

***x6.18.** There exists exactly one pair of partial functions

$$\alpha : N \times N \rightarrow Z, \quad \beta : N \times N \rightarrow Z,$$

with common domain of definition $\{(n, m) \mid n, m \neq 0\}$ which satisfy the following identities for all $n, m, k > 0$:

$$\begin{aligned} n \neq m &\implies \alpha(n, m) = \beta(m, n), \\ \alpha(n, n) &= 1, \\ \beta(n, n) &= 0, \\ \alpha(n, n+k) &= \alpha(n, k) - \beta(n, k), \\ \beta(n, n+k) &= \beta(n, k). \end{aligned}$$

It follows that for any two natural numbers $n, m \neq 0$,

$$\gcd(n, m) = \alpha(n, m)n + \beta(n, m)m.$$

Explain the algorithmic significance of the theorem, with examples of computations of $\alpha(n, m)$ and $\beta(n, m)$.

6.38. Definition. For each finite partial function $g : A \rightarrow E$, the **neighborhood** determined by g in the poset $(A \rightarrow E)$ is the set

$$N(g) =_{\text{df}} \{f : A \rightarrow E \mid g \subseteq f\}$$

of all extensions of g . A set $G \subseteq (A \rightarrow E)$ is **open in the topology of pointwise convergence** if

$$f \in G \implies (\exists g, \text{finite})[f \in N(g) \subseteq G].$$

x6.19. Prove that the family of open sets in $(A \rightarrow E)$ defined in **6.38** is a topology by **4.26**, and a mapping

$$\pi : (A \rightarrow E) \rightarrow (B \rightarrow M)$$

is continuous in this topology by **6.24** if and only if it is continuous by **6.22**.

6.39. Definition. A subset $G \subseteq P$ of an inductive poset is **Scott open** if (1) it is upward closed, i.e.

$$x < y \ \& \ x \in G \implies y \in G,$$

and (2) for every non-empty chain $S \subseteq P$,

$$\sup S \in G \implies (\exists x \in S)[x \in G].$$

***x6.20.** Prove that the family of Scott open subsets of an inductive poset P is a topology. **HINT:** Notice that a set $F \subseteq P$ is Scott *closed* if it is downward closed and for every non-empty chain S , $S \subseteq F \implies \sup S \in F$, and work with the closed sets rather than the open ones. Instead of showing that the intersection $G_1 \cap G_2$ of open sets is open, it is somewhat easier to show that the union of closed sets is closed.

***x6.21.** Suppose P and Q are inductive posets and $\pi : P \rightarrow Q$ is a mapping. Prove that π is continuous in the relevant Scott topologies if and only if π is monotone and for every non-empty chain $S \subseteq P$,

$$\pi(\sup S) = \sup \pi[S].$$

HINT: Prove and use the fact that for every $c \in P$, the set $\{x \in P \mid x \leq c\}$ is Scott closed.

***x6.22.** Suppose A is a countable set and $\pi : (A \rightarrow E) \rightarrow (B \rightarrow M)$ is a mapping. Show that π is continuous (by the definition in **6.22**) if and only if it is continuous with respect to the Scott topologies in the posets $(A \rightarrow E)$ and $(B \rightarrow M)$.

The Continuous Least Fixed Point Theorem is often formulated for the class of *directed-complete* posets, particularly in Computer Science texts.

6.40. Definition. A subset $S \subseteq P$ of a poset P is **directed** if any two members of S have an upper bound in S ,

$$x, y \in S \implies (\exists z \in S)[x \leq z \ \& \ y \leq z].$$

A poset P is **directed-complete** (a **dcpo**) if every directed $S \subseteq P$ has a least upper bound.

x6.23. Every chain in a poset is a directed set, hence, every dcpo is an inductive poset and the least fixed point theorems hold for directed-complete posets.

x6.24. For each A and E , the posets $(A \rightarrow E)$ and $(A \multimap E)$ are directed-complete.

x6.25. A mapping $\pi : (A \rightarrow E) \rightarrow (B \rightarrow M)$ is continuous if and only if for each directed $S \subseteq (A \rightarrow E)$,

$$\pi(\sup S) = \sup \pi[S].$$

x6.26. The product $P_1 \times P_2$ (Problem **x6.3**) of two directed-complete partial orderings is also directed-complete.

***x6.27.** Every countable, inductive poset is directed-complete.

Actually the notions of *inductive* and *directed-complete* are equivalent; for a monotone mapping $\pi : P \rightarrow Q$ on one inductive poset to another, the equation

$$\pi(\sup S) = \sup \pi[S] \tag{6.25}$$

holds for all non-empty chains $S \subseteq P$ if and only if it holds for all non-empty directed sets $S \subseteq P$; and the characterization of Scott continuity in Problem ***x6.22** holds whether A is countable or not. The proofs of these results are not elementary and require the Axiom of Choice; see Problems ***x9.20 -x9.23**.

Chapter 7

WELL ORDERED SETS

7.1. A well ordered set¹ is a poset

$$U = (Field(U), \leq_U),$$

where $\leq_U$ is a wellordering on $Field(U)$, i.e. a linear (total) ordering on $Field(U)$ such that every non-empty $X \subseteq Field(U)$ has a least member. Associated with U is also its **strict ordering** $<_U$,

$$x < y \iff x <_U y \iff_{\text{df}} x \leq_U y \ \& \ x \neq y.$$

As we did with arbitrary posets in the preceding chapter, we will usually identify U with its field, talk about *the points* or *subsets* of U , meaning the members and subsets of $Field(U)$, etc.

The most basic results of the last two chapters were all proved by some combination of the coupled techniques

definition by recursion - proof by induction. (7.1)

In the simplest case, some function $f : N \rightarrow E$ is defined by recursion, some properties of f are proved by induction and these in turn imply the theorem we want. Typical are the Continuous Least Fixed Point and the Schröder-Bernstein theorems which say nothing (explicitly) about recursion, induction or any functions with domain N , but whose proofs most assuredly use precisely these notions. We based the proof of the Recursion Theorem 5.6 directly on the Induction Axiom for the natural numbers. The key fact, however, which can be generalized is that N is well ordered by its natural ordering. Here we will generalize 5.6 to a powerful TRANSFINITE RECURSION THEOREM 7.24 which justifies *definition by recursion* of functions $f : U \rightarrow E$ on every well ordered set U . Coupled with HARTOGS' THEOREM 7.34 which guarantees the existence of "arbitrarily large" well ordered sets, this makes it possible to apply the basic idea of (7.1) in situations far removed from the natural numbers. A typical application is the

¹Do we dare call them *wosets*? It's not much worse than *posets* and it would sure save a lot of key strokes.



Figure 7.1. The low end of a long wellordering.

FIXED POINT THEOREM 7.35 and its corollary, the **LEAST FIXED POINT THEOREM 7.36**, which is just **6.21** without the countable continuity hypothesis.

7.2. A set A is **well orderable** if it admits a wellordering, so it is the field of some well ordered set $(A, \leq)$. One of the chief lessons of this chapter is that well orderable sets behave much better than arbitrary sets, for example any two of them are comparable in cardinality, either $A \leq_c B$ or $B \leq_c A$. In fact, *every set is well orderable*. Zermelo showed this in 1904, settling with one brilliant stroke the problem of Cardinal Comparability and a whole slew of related, regularity questions about arbitrary sets. We will prove Zermelo's Wellordering Theorem in the next chapter, after we introduce the Axiom of Choice on which it is based. It is worth pointing out here, however, that the mathematical content of this fundamental result is just the sum of the Transfinite Recursion and Hartogs' Theorems: the Axiom of Choice simply allows us to put the two together.

7.3. Exercise. If C is well orderable and $A \leq_c C$, then A is well orderable.

7.4. Exercise. If C is well orderable and there exists a surjection $f : C \rightarrow A$, then $A \leq_c C$, and hence A is also well orderable.

7.5. Successor and limit points. The set N of natural numbers is well ordered by its natural ordering, and so is each of its finite initial segments

$$[0, n) = \{i \in N \mid i < n\}.$$

Every well ordered set U looks at its low end like an initial segment of N . If it is not empty, it must have a least member which is typically denoted by 0 rather than $\perp$,

$$0 = 0_U =_{\text{df}} \text{the least element of } U. \quad (7.2)$$

It is pictured by a square in Figure 7.1. Each $x \in U$ other than the maximum (which may or may not exist) has an element following it immediately,

$$S(x) = S_U(x) =_{\text{df}} \inf_U \{y \in U \mid x < y\}. \quad (7.3)$$

The values of the partial function $S : U \rightarrow U$ are the **successor points** of U . In addition, U may have **limit points** which are above 0 but not the successor of anything:

$$\text{Limit}_U(x) \iff_{\text{df}} 0 < x \ \& \ (\forall u < x)(\exists v)[u < v < x]. \quad (7.4)$$

These are pictured by black boxes in Figure 7.1. The first limit point of U is typically denoted by

$$\omega = \omega_U =_{\text{df}} \inf\{x \in U \mid \text{Limit}(x)\}, \quad (7.5)$$

when it exists, the points below it are the **finite points** and the points above it (including ω) are the **infinite points** of U . If U is infinite, then the function $\pi : N \rightarrow U$ defined by the recursion

$$\begin{aligned} \pi(0) &= 0_U = \text{the least member of } U, \\ \pi(n+1) &= S_U(\pi(n)), \end{aligned} \quad (7.6)$$

establishes an order-preserving correspondence of N with the finite points of U .

7.6. Exercise. For each subset $I \subseteq U$ of a well ordered set U , the restriction

$$x \leq_I y \iff_{\text{df}} x \leq_U y \ \& \ x, y \in I$$

of $\leq_U$ to I is a wellordering, so that I is a well ordered set in its own right with this ordering.

7.7. Definition. A well ordered set U is an **initial segment** of V if $\text{Field}(U)$ is a downward closed subset of $\text{Field}(V)$ and $\leq_U$ is the restriction of $\leq_V$ to $\text{Field}(U)$:

$$\begin{aligned} U \sqsubseteq V &\iff_{\text{df}} \text{Field}(U) \subseteq \text{Field}(V) \\ &\& (\forall x, y \in \text{Field}(U))[x \leq_U y \iff x \leq_V y] \\ &\& (\forall y \in \text{Field}(U))(\forall x \leq_V y)[x \in \text{Field}(U)]. \end{aligned} \quad (7.7)$$

Clearly V is an initial segment of itself, the **trivial** one. With each $x \in V$ we associate the **proper initial segment** of points strictly below x ,

$$\text{seg}(y) = \text{seg}_V(y) =_{\text{df}} \{x \in V \mid x <_V y\} \subsetneq U. \quad (7.8)$$

More precisely, this is the field of $\text{seg}(y)$, but the ordering is determined by V and we will talk about initial segments as if they were just sets, as usual.

7.8. Exercise. If 0 is the least element of U , then $\text{seg}(0) = \emptyset$, and if $x \in U$ has a successor, then

$$\text{seg}(S(x)) = \text{seg}(x) \cup \{x\}.$$

7.9. Proposition. A set I is an initial segment of a well ordered set U if and only if $I = U$ or for some $x \in U$, $I = \text{seg}(x)$.

Proof. If $I \subsetneq U$, let $x = \inf(U \setminus I)$ so that immediately,

$$y \in \mathbf{seg}(x) \implies y < x \implies y \in I,$$

and it is enough to prove

$$y \in I \implies y < x$$

to verify that $I = \mathbf{seg}(x)$. Towards a contradiction, if $y \in I$ but $y \not< x$, then we must have $x \leq y$, which implies $x \in I$ because I is downwards closed, contradicting the choice of x . The converse is trivial. $\dashv$

7.10. Exercise. *The family of initial segments of a well ordered set U is well ordered by the relation $\subseteq$.*

The general idea is to view a well ordered set U as a generalization of the natural number sequence $0, 1, 2, \dots$, possibly shorter than or of equal length to N , typically much longer. The particular members of U will be of little consequence, it is the *length* of the sequence in which we will be interested. We introduce here the general notion of isomorphism which relates posets with the same *shape*, the shape of a well ordered set being just a “length.”

7.11. Definition. *A function $\pi : P \rightarrow Q$ from one poset into another is **order-preserving** if for all $x, y \in P$,*

$$x \leq_P y \iff \pi(x) \leq_Q \pi(y);$$

*a **similarity** is an order-preserving bijection $\pi : P \rightarrow Q$, and if one such exists we call P and Q **similar**, **order isomorphic** or **copies** of each other, and we write*

$$P =_o Q \iff (\exists \pi : P \rightarrow Q)[\pi \text{ a similarity}].$$

The subscript o in $=_o$ stands for “order type,” a fancier expression for “shape.” Notice that by our general convention of talking about a poset as if it were its field, we write $\pi : P \rightarrow Q$ for similarities instead of the more explicit $\pi : \mathbf{Field}(P) \rightarrow \mathbf{Field}(Q)$.

7.12. Exercise. *Every order-preserving $\pi : P \rightarrow Q$ from one poset to another is monotone; but there exist monotone mappings which are not order-preserving.*

7.13. Exercise. *If P and Q are linear posets, then a function $f : P \rightarrow Q$ is order-preserving if and only if it is **strictly monotone**, i.e. $x <_P y \implies f(x) <_Q f(y)$. In particular, order-preserving functions on well ordered sets are strictly monotone, and hence one-to-one.*

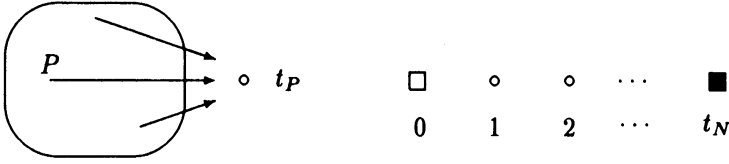


Figure 7.2. The successor poset to P and $\text{Succ}(N)$.

7.14. Exercise. For all posets P, Q, R ,

$$\begin{aligned} P &=_o P, \\ P =_o Q &\implies Q =_o P, \\ P =_o Q \ \& \ Q =_o R &\implies P =_o R. \end{aligned}$$

7.15. Lemma. If a poset P is similar to a well ordered set U , then it is also well ordered.

Proof. Given $\emptyset \neq X \subseteq P$, let $p \in U$ be the $\leq_U$ -least element of the image $\pi[X]$ and verify (easily) that $x = \pi^{-1}(p)$ is $\leq_P$ -least in X , because π preserves the orderings. $\dashv$

We can construct explicitly some fairly long wellorderings by starting with N and its finite initial segments and applying repeatedly several natural operations on posets which yield well ordered sets on well ordered arguments. Here we look at just one of these, leaving the rest for the problems.

7.16. The **successor** of a poset P is obtained by adding a new point above all the members of P . To be specific, we can choose to add to the field of P the object

$$t_P =_{\text{df}} \mathbf{r}(\text{Field}(P)) \quad (7.9)$$

which is guaranteed by **3.11** to be a new element, and we set

$$x \leq_{\text{Succ}(P)} y \iff_{\text{df}} x \leq_P y \vee [x \in P \ \& \ y = t_P] \vee x = y = t_P. \quad (7.10)$$

If P is finite with n elements, then $\text{Succ}(P)$ has $n + 1$ elements, in fact, easily $\text{Succ}([0, n)) =_o [0, (n + 1))$. On the other hand, $\text{Succ}(N)$ is countably infinite, but with a different, “longer” ordering than N , it has a maximum element which comes after all the natural numbers.

7.17. Exercise. If $P =_o Q$, then $\text{Succ}(P) =_o \text{Succ}(Q)$.

7.18. Exercise. If U is well ordered, so is $\text{Succ}(U)$.

Using this successor operation on posets, we can view each well ordered set U as a proper initial segment of another,

$$U = \mathbf{seg}_{\text{Succ}(U)}(t_U) \subsetneq \text{Succ}(U). \quad (7.11)$$

7.19. Definition. A mapping $\pi : P \rightarrow P$ on a poset to itself is **expansive**, if for all $x \in P$, $x \leq \pi(x)$.

7.20. Theorem. Every order-preserving injection $\pi : U \rightarrow U$ of a well ordered set into itself is expansive.

Proof. Towards a contradiction, assume that $\pi : U \rightarrow U$ is order-preserving but that for some $x \in U$, $\pi(x) < x$, and let

$$x^* = \inf\{x \in U \mid \pi(x) < x\}.$$

Thus, $\pi(x^*) < x^*$, so $\pi(\pi(x^*)) < \pi(x^*)$ since π is an order-preserving injection, and this contradicts the choice of x^* . $\neg$

7.21. Corollary. No well ordered set is similar with one of its proper initial segments, and hence no two distinct initial segments of a well ordered set are similar.

Proof. Every similarity $\pi : U \rightarrow \mathbf{seg}(x)$ is (in particular) an order-preserving injection of U into U , so we cannot have $\pi(x) < x$, by the theorem. $\neg$

Because a well ordered set may have limit points in addition to its 0 and its successor points, it is easiest to generalize the principles of proof by *complete induction* and definition by *complete recursion*.

7.22. Transfinite Induction Theorem. For every well ordered set U and every unary definite condition P ,

$$(\forall y \in U)[(\forall x < y)P(x) \implies P(y)] \implies (\forall y \in U)P(y).$$

Proof. Assuming the opposite, towards a contradiction, let

$$y^* =_{\text{df}} \inf\{y \in U \mid (\forall x < y)P(x) \ \& \ \neg P(y)\};$$

the hypothesis yields $P(y^*)$, which contradicts the choice of y^* . $\neg$

In specific cases, it is often just as easy to prove $(\forall y \in U)P(y)$ by contradiction rather than appeal to **7.22**, in effect repeating this little argument. It depends on the statement to be proved and how much one is annoyed by

dealing with negative statements. We will illustrate both styles. Incidentally, the term “transfinite” is used because U may be longer than N , the theorem also holds, of course, when U is finite or similar with N .

The next lemma is the key step in the proof of the fundamental theorem which follows it.

7.23. Lemma. *Suppose U is a well ordered set and $h : (U \rightarrow E) \rightarrow E$ maps the partial functions from U to E into E . It follows that for every $t \in U$, there exists exactly one function*

$$\sigma_t : \mathbf{seg}(t) \rightarrow E$$

which satisfies the identity

$$\sigma_t(x) = h(\sigma_t \upharpoonright \mathbf{seg}(x)) \quad (x < t). \quad (7.12)$$

Proof. By Transfinite Induction, assume that for each $u < t$ there exists exactly one function $\sigma_u : \mathbf{seg}(u) \rightarrow E$ such that

$$\sigma_u(x) = h(\sigma_u \upharpoonright \mathbf{seg}(x)) \quad (x < u). \quad (7.13)$$

The induction hypothesis gives us nothing if $t = 0_U$ is the least point in U , but the required conclusion is trivial in this case taking $\sigma_0 = \emptyset$. If $u = Sv$ is a successor point in U , we set

$$\sigma_u = \sigma_v \cup \{(v, h(\sigma_v))\};$$

now (7.13) holds for $x < v$ by the induction hypothesis and it holds for $x = v$ by the definition. For the last case, when t is limit, we need a

Lemma. *The set of functions $\{\sigma_u \mid u < t\}$ is a chain under $\subseteq$, i.e.*

$$x < u < v < t \implies \sigma_u(x) = \sigma_v(x). \quad (7.14)$$

Proof. Assume not and let x be least such that (7.14) fails for some $u < v < t$. This means that

$$\sigma_u \upharpoonright \mathbf{seg}(x) = \sigma_v \upharpoonright \mathbf{seg}(x),$$

and then by the identity which σ_u, σ_v satisfy,

$$\begin{aligned} \sigma_u(x) &= h(\sigma_u \upharpoonright \mathbf{seg}(x)) \\ &= h(\sigma_v \upharpoonright \mathbf{seg}(x)) \\ &= \sigma_v(x), \end{aligned}$$

which contradicts the choice of x .

We now take

$$\sigma_t = \bigcup \{\sigma_u \mid u < t\};$$

this is a function with domain $\mathbf{seg}(t)$ by the Lemma, and it satisfies (7.12), since for each $x < t$,

$$\begin{aligned} \sigma_t(x) &= \sigma_u(x) && \text{for some } u \text{ such that } x < u < t, \\ &= h(\sigma_u \upharpoonright \mathbf{seg}(x)) && \text{by ind. hyp.,} \\ &= h(\sigma_t \upharpoonright \mathbf{seg}(x)) && \text{since } \sigma_u \upharpoonright \mathbf{seg}(x) = \sigma_t \upharpoonright \mathbf{seg}(x). \end{aligned}$$

This completes the proof of existence of σ_t , and its uniqueness is obvious from (7.12). $\dashv$

7.24. Transfinite Recursion Theorem. *For each well ordered set U and each function $h : (U \rightarrow E) \rightarrow E$, there exists exactly one function $f : U \rightarrow E$ which satisfies the identity*

$$f(x) = h(f \upharpoonright \mathbf{seg}(x)) \quad (x \in U). \quad (7.15)$$

Proof. Consider the next well ordered set $\text{Succ}(U)$ to U which has some point $t = t_U$ on top of U , and the extension $h' : (\text{Succ}(U) \rightarrow E) \rightarrow E$ of h defined by

$$h'(\sigma) = \begin{cases} h(\sigma), & \text{if } \text{Domain}(\sigma) \subseteq U, \\ e^*, & \text{if } t \in \text{Domain}(\sigma), \end{cases}$$

where e^* is some arbitrary member of E , of no consequence. The function h' has the correct domain for applying the lemma to $\text{Succ}(U)$ and h' , because $\mathbf{seg}_{\text{Succ}(U)}(t) = U$. For the top point t the Lemma gives a unique $f = \sigma_t : U \rightarrow E$ which satisfies (7.12) for all $x \in U$. $\dashv$

Perhaps the simplest, non-trivial application of Transfinite Recursion is the definition of transfinite orbits for mappings of an inductive poset into itself. We consider first the basic case of *expansive* mappings, defined in 7.19. Expansive mappings are related to the monotone mappings we studied in the last chapter, but the two notions do not coincide; for example, the constant mapping $X \mapsto \emptyset$ on $\mathcal{P}(N)$ is obviously monotone but not expansive, while

$$\pi(X) = \begin{cases} X \cup \{1\} & \text{if } 0 \in X, \\ X \cup \{2\} & \text{if } 0 \notin X \end{cases}$$

is expansive but (easily) not monotone. It turns out, however, that results about expansive mappings can often be translated into similar results about monotone mappings.

7.25. Iteration Lemma. *Suppose $\pi : P \rightarrow P$ is an expansive mapping on an inductive poset and U is a well ordered set. There exists a unique*

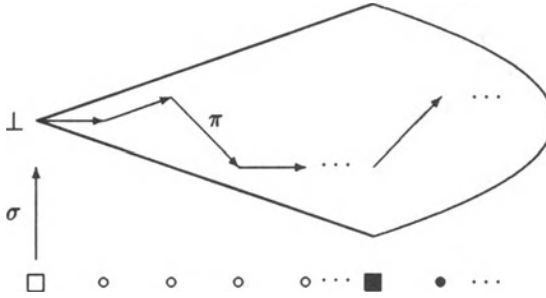


Figure 7.3. The transfinite orbit of an expansive mapping.

function $\sigma : U \rightarrow P$ which satisfies the following conditions:

$$\begin{aligned} \sigma(0) &= \perp, \\ \text{if } y = S(x), \text{ then } \sigma(y) &= \pi(\sigma(x)), \\ \text{if } \text{Limit}(y), \text{ then } \sigma(y) &= \sup_P \{\sigma(x) \mid x < y\}. \end{aligned} \quad (7.16)$$

In addition, this σ is monotone from U to P , i.e.

$$x \leq y \implies \sigma(x) \leq_P \sigma(y). \quad (7.17)$$

Proof. The conditions in (7.16) just about give a definition of σ by transfinite recursion, except that there is a problem in the limit case if the set $\{\sigma(x) \mid x < y\}$ is not a chain in P . To account for this possibility, we define σ by appealing to 7.24 so that it satisfies the following:

$$\sigma(y) = \begin{cases} \perp, & \text{if } y = 0, \\ \pi(\sigma(x)), & \text{if } y = S(x) \text{ for some } x, \\ \sup_P \{\sigma(x) \mid x < y\}, & \text{if } \text{Limit}(y) \\ & \& (\forall x_1 < x_2 < y)[\sigma(x_1) \leq_P \sigma(x_2)], \\ \perp, & \text{otherwise,} \end{cases}$$

where $\leq_U$ is the wellordering of U , as in the statement of the theorem. The result follows directly from the following lemma, which implies in particular that the “otherwise” case in the definition of σ never comes up.

Lemma. For each $y \in U$,

$$x_1 < x_2 \leq y \implies \sigma(x_1) \leq_P \sigma(x_2). \quad (7.18)$$

Proof. Assume not and let y be least in U such that (7.18) fails. Since (7.18) holds vacuously when $y = 0$ is the least element in U , we need consider only two cases.

CASE 1. $y = S(x)$ is a successor point. Assume $x_1 < x_2 \leq y$. If $x_2 \leq x$, we get $\sigma(x_1) \leq_P \sigma(x_2)$ by the choice of y . The only other possibility is that $x_2 = y$, but then $\sigma(x_1) \leq_P \sigma(x)$ by the choice of y , and $\sigma(x) \leq \pi(\sigma(x)) = \sigma(y)$ by the expansiveness of π . Thus, (7.18) holds for y , which is a contradiction.

CASE 2. y is limit. By the choice of y ,

$$x_1 < x_2 < y \implies \sigma(x_1) \leq \sigma(x_2), \quad (7.19)$$

so in order to get a contradiction, we only need show that $x_1 < y \implies \sigma(x_1) \leq_P \sigma(y)$. This holds because (7.19) also implies immediately that $\sigma(y) = \sup_P \{\sigma(x) \mid x < y\}$. $\dashv$

The transfinite orbit $\sigma : U \rightarrow P$ of a mapping $\pi : P \rightarrow P$ guaranteed by the Iteration Lemma is obviously an extension of the orbit $(n \mapsto x_n)$ which we defined in the proof of the Continuous Least Fixed Point Theorem **6.21**, at least if the well ordered set U is longer than N . It is one of the tools we will use in the proof of the Least Fixed Point Theorem, as follows.

7.26. Plan for a proof. Suppose that for the given inductive poset P , we can construct a well ordered set U such that *there exists no injection* $\sigma : U \rightarrow P$. In particular, the transfinite orbit $\sigma : U \rightarrow P$ of **7.25** cannot be an injection, and there exist $x < y$ such that $\sigma(x) = \sigma(y)$. The monotonicity of σ implies that

$$x \leq u \leq y \implies \sigma(x) = \sigma(u),$$

x has a successor since it is not maximum in U , $x < Sx \leq y$ and, hence,

$$\sigma(x) = \sigma(Sx) = \pi(\sigma(x));$$

in other words, the point $\sigma(x)$ is a fixed point of π . Thus, to prove that every expansive mapping $\pi : P \rightarrow P$ on an inductive poset has a fixed point, it is sufficient to show that *for each set P , there exists some well ordered set U which cannot be injected into P* . This is precisely Hartogs' Theorem, for which we aim next. To show it, we must study in some detail the question of comparability of well ordered sets as to length.

The picture of the typical well ordered set in Figure 7.1 suggests that we should be able to compare any two of them, line them up side-by-side, the least element 0_U of one facing the least element 0_V of the other, the next $S_U(0_U)$ facing $S_V(0_V)$, the first limit point ω_U (if it exists) facing ω_V , etc. until we run out of elements in either U or V . The precise version of this fact is a generalization of the Uniqueness Theorem for the natural numbers **5.4**.

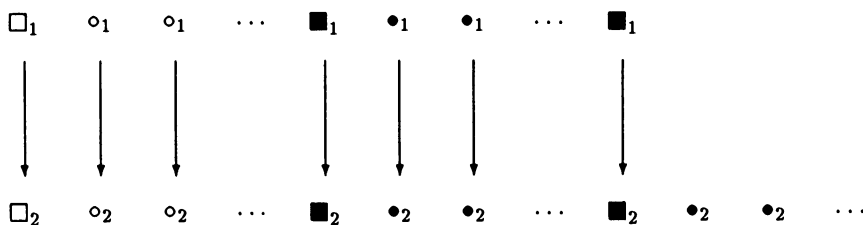


Figure 7.4. Portrait of an initial similarity.

7.27. Definition. *An initial similarity*

$$\pi : U \rightarrow \pi[U] \subseteq V$$

from one well ordered set into another is a similarity of U with an initial segment of V . If such an initial similarity exists, we say that U is **less than or equal to V in length**, in symbols:

$$U \leq_o V \iff (\exists I \subseteq V)[U =_o I]. \quad (7.20)$$

We also write

$$U <_o V \iff_{\text{df}} U \leq_o V \ \& \ U \neq_o V. \quad (7.21)$$

By 7.9, every initial similarity $\pi : U \rightarrow V$ is either a similarity with V , or one with a proper initial segment of V , so that

$$U <_o V \iff (\exists x \in V)[U =_o \text{seg}_V(x)]. \quad (7.22)$$

7.28. Exercise. If $\pi : U \rightarrow V$ and $\rho : V \rightarrow W$ are initial similarities, then so is their composition $\rho\pi : U \rightarrow W$.

7.29. Proposition. For all well ordered sets U, V, W ,

$$\begin{aligned} U &\leq_o U, \\ U \leq_o V \ \& \ V \leq_o W &\implies U \leq_o W, \\ U \leq_o V \ \& \ V \leq_o U &\implies U =_o V. \end{aligned}$$

Proof. Only the third of these assertions needs proof and it follows from 7.21. The composition $\rho\pi$ of the initial similarities $\pi : U \rightarrow V$, $\rho : V \rightarrow W$ witnessing the hypothesis is an initial similarity $\rho\pi : U \rightarrow W$, which if it were not onto, would witness that U is similar with one of its proper initial segments; so it is a bijection, and then π must also be a bijection. $\dashv$

7.30. Theorem. A function $\pi : U \rightarrow V$ is an initial similarity of a well ordered set into another if and only if it satisfies the identity

$$\pi(x) = \inf_V \{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\}. \quad (7.23)$$

Proof. If $\pi : U \rightarrow V$ is an initial similarity, then it is order-preserving and one-to-one, so it satisfies

$$(\forall u <_U x)[\pi(u) <_V \pi(x)], \quad (7.24)$$

and hence

$$z = \inf_V \{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\} \leq_V \pi(x).$$

Since π is initial and $z \leq_V \pi(x)$, there exists some $u \in U$ such that $\pi(u) = z$. Assuming towards a contradiction that $z = \pi(u) <_V \pi(x)$, we infer that $u <_U x$ because π is an order-preserving injection, and hence $\pi(u) <_V z$ by the definition of z , which is absurd since $z = \pi(u)$.

Conversely, if $\pi : U \rightarrow V$ satisfies (7.23), then it is an order-preserving injection, since by (7.23), $u <_U x \implies \pi(u) <_V \pi(x)$. Suppose the image $\pi[U]$ is not an initial segment of V and choose x least in U such that there exists some $y <_V \pi(x)$, $y \notin \pi[U]$. Now $\pi[\text{seg}_U(x)] \sqsubseteq V$ by the choice of x ; it is a proper initial segment since it does not contain y ; so $\pi[\text{seg}_U(x)] = \text{seg}_V(z)$ for some $z \in V$, and (7.23) yields $\pi(x) = z$. Thus $y <_V z$ and $y \in \text{seg}_V(z) = \pi[\text{seg}_U(x)]$, which is absurd. $\dashv$

7.31. Comparability Theorem for Well Ordered Sets. *For any two well ordered sets U, V , either $U \leq_o V$ or $V \leq_o U$.*

Proof. The result is trivial if $V = \emptyset$, so we may assume the minimum 0_V exists. By the Transfinite Recursion Theorem 7.24, there exists a function $\pi : U \rightarrow V$ which satisfies the identity

$$\pi(x) = \begin{cases} \inf_V \{y \in V \mid (\forall u <_U x)[\pi(u) <_V y]\}, & \text{if } (\exists y \in V)(\forall u <_U x)[\pi(u) <_V y], \\ 0_V, & \text{otherwise.} \end{cases} \quad (7.25)$$

In pedantic detail, we are applying here 7.24 with the mapping $h : (U \rightarrow E) \rightarrow E$, defined by

$$h(p) = \begin{cases} \inf_V \{y \in V \mid (\forall u \in \text{Domain}(p))[p(u) <_V y]\}, & \text{if } (\exists y \in V)(\forall u \in \text{Domain}(p))[p(u) <_V y], \\ 0_V, & \text{otherwise.} \end{cases}$$

We now distinguish two possibilities.

CASE 1. *For every $x \neq 0_U$, $\pi(x) \neq 0_V$.* This means that the second case in (7.25) never applies, π satisfies the identity (7.23) and it must be an initial similarity by 7.30.

CASE 2. *For some $a \in U$, $a \neq 0_U$, we have $\pi(a) = 0_V$.* Let a be least in U , $a \neq 0_U$ and such that $\pi(a) = 0_V$, and consider the restriction

$$\rho = (\pi \upharpoonright \text{seg}_U(a)) : \text{seg}_U(a) \rightarrow V.$$

Now ρ satisfies (7.23), so by 7.30 it is an initial similarity of $\text{seg}_U(a)$ into V . In particular, the image $\rho[\text{seg}_U(a)] = \pi[\text{seg}_U(a)]$ is an initial segment of V ; if it were proper, then $\pi[\text{seg}_U(a)] = \text{seg}_V(z)$ for some $z \in V$ and (7.25) would yield $\pi(a) = z \neq 0_V$, contradicting the choice of a ; hence, $\pi[\text{seg}_U(a)] = V$. Thus, $V =_o \text{seg}_U(a)$, which gives us an initial similarity of V into U . $\dashv$

This fundamental theorem has a host of corollaries, some of which are worth listing immediately. This first gives us an easier way to compare well ordered sets.

7.32. Corollary. *For all well ordered sets U, V ,*

$$U \leq_o V \iff (\exists \pi : U \rightarrow V)[\pi \text{ is order-preserving}].$$

Proof. Suppose $\pi : U \rightarrow V$ is order-preserving but $U \not\leq_o V$, so that $V <_o U$. It follows that $V =_o \text{seg}_U(x)$ for some x by (7.22), and composing the order-preserving injections we get an injection $\rho : U \rightarrow \text{seg}_U(x)$ which is still order-preserving and violates 7.20. $\dashv$

7.33. Corollary. Wellfoundedness of $\leq_o$. *Every non-empty class $\mathcal{E}$ of well ordered sets has a $\leq_o$ -least member, i.e. for some $U_0 \in \mathcal{E}$ and all $U \in \mathcal{E}$, $U_0 \leq_o U$.*

Proof. The hypothesis gives us some $W \in \mathcal{E}$, and if W is $\leq_o$ -least in $\mathcal{E}$, there is nothing to prove. If not, then 7.31 implies that there exists well ordered sets in $\mathcal{E}$ which are similar with proper initial segments of W , so the set

$$J =_{\text{df}} \{x \in W \mid (\exists U \in \mathcal{E})[U =_o \text{seg}_W(x)]\} \quad (7.26)$$

is non-empty and it has a $\leq_W$ -least element x . By the definition of J , there exist some $U_0 \in \mathcal{E}$ such that $U_0 =_o \text{seg}_W(x)$ and we claim that this U_0 is $\leq_o$ -least in $\mathcal{E}$. To prove it, assume towards a contradiction that for some $U \in \mathcal{E}$, $U_0 \not\leq_o U$; hence $U <_o U_0 =_o \text{seg}_W(x)$; hence $U =_o \text{seg}_W(y)$ for some $y <_W x$, contradicting the choice of x . $\dashv$

Most often this is applied when $\mathcal{E}$ is actually a set, a family of well ordered sets, but occasionally it is convenient to cite it more generally for classes. For example, there exists a $\leq_o$ -least well ordered set which has a limit point, this would be $\text{Succ}(N)$.

After all this work, still we have not constructed any uncountable well ordered sets and it might appear that all our results apply only to peculiar, long reshufflings of N . Next comes the second basic theorem of the chapter which rectifies the situation.

7.34. Hartogs' Theorem. *There is a definite operation $\chi(A)$ which associates with each set A , a well ordered set*

$$\chi(A) = (h(A), \leq_{\chi(A)}),$$

such that $\chi(A) \not\leq_c A$: i.e. there exists no injection $\pi : h(A) \rightarrow A$. Moreover, $\chi(A)$ is $\leq_o$ -minimal with this property, i.e. for every well ordered set W ,

$$W \not\leq_c A \implies \chi(A) \leq_o W. \quad (7.27)$$

Proof. First set

$$WO(A) =_{\text{df}} \{U \mid U = (Field(U), \leq_U) \quad (7.28)$$

is a well ordered set with $Field(U) \subseteq A\}$,

and let $\sim_A$ be the restriction of the definite condition $=_o$ to $WO(A)$,

$$U \sim_A V \iff_{\text{df}} U, V \in WO(A) \text{ \& } U =_o V.$$

Clearly $\sim_A$ is an equivalence relation on $WO(A)$, and we set

$$h(A) =_{\text{df}} \llbracket WO(A)/\sim_A \rrbracket \subseteq \mathcal{P}(WO(A)). \quad (7.29)$$

We order the equivalence classes in $h(A)$ by their “representatives,”

$$[U/\sim_A] \leq_{\chi(A)} [V/\sim_A] \iff_{\text{df}} U \leq_o V; \quad (7.30)$$

this makes sense because if

$$[U/\sim_A] = [U'/\sim_A], [V/\sim_A] = [V'/\sim_A], \text{ and } U \leq_o V,$$

then $U' =_o U \leq_o V =_o V'$. The fact that $\leq_{\chi(A)}$ is a wellordering of $h(A)$ follows easily from the general properties of $\leq_o$, **7.31** and **7.33**. Taking the negation of both sides of (7.30) we infer its strict version,

$$V <_o U \iff [V/\sim_A] <_{\chi(A)} [U/\sim_A] \quad (U, V \in WO(A)). \quad (7.31)$$

The basic properties of the Hartogs operation are embodied in the following

Lemma. *For every $\alpha = [U/\sim_A] \in h(A)$,*

$$\text{seg}_{\chi(A)}(\alpha) = \{[\text{seg}_U(x)/\sim_A] \mid x \in U\} =_o U.$$

In particular, every proper initial segment of $\chi(A)$ is similar with some $U \in WO(A)$, and every $U \in WO(A)$ is similar with a proper, initial segment of $\chi(A)$.

Proof. We verify first the identity

$$\text{seg}_{\chi(A)}(\alpha) = \{[\text{seg}_U(x)/\sim_A] \mid x \in U\}.$$

If $\beta = [V/\sim_A] <_{\chi(A)} \alpha$, then $V <_o U$ from (7.31), and hence $V =_o \text{seg}_U(x)$ for some $x \in U$, so that $\beta = [\text{seg}_U(x)/\sim_A]$. Conversely, for each $x \in U$, $\text{seg}_U(x) <_o U$, hence, $[\text{seg}_U(x)/\sim_A] <_{\chi(A)} [U/\sim_A] = \alpha$, again by (7.31).

To show the similarity

$$U =_o \text{seg}_{\chi(A)}(\alpha) = \{[\text{seg}_U(x)/\sim_A] \mid x \in U\},$$

define $\rho : U \rightarrow h(A)$ by

$$\rho(x) = [\text{seg}_U(x)/\sim_A], \quad (x \in U);$$

ρ is a similarity of U with the image $\rho[U]$, because

$$\begin{aligned} x <_U y &\iff \text{seg}_U(x) \subsetneq \text{seg}_U(y) &\iff \text{seg}_U(x) <_o \text{seg}_U(y) \\ & &\iff [\text{seg}_U(x)/\sim_A] <_o [\text{seg}_U(y)/\sim_A]. \end{aligned}$$

Suppose now, towards a contradiction, that there exists an injection

$$\pi : h(A) \rightarrow A,$$

and let $B = \pi[h(A)] \subseteq A$ be its image. The injection π copies the wellordering of $h(A)$ to a wellordering of B ,

$$x \leq_B y \iff_{\text{df}} \pi^{-1}(x) \leq_{\chi(A)} \pi^{-1}(y) \quad (x, y \in B),$$

so that $U = (B, \leq_B)$ is a well ordered subset of A , and by its definition,

$$U =_o \chi(A). \tag{7.32}$$

But U is similar with a proper initial segment of $\chi(A)$ by the Lemma, and hence $U <_o \chi(A)$, which contradicts $U =_o \chi(A)$.

To show the minimality of $\chi(A)$, notice that if $W <_o \chi(A)$, then $W =_o \text{seg}_{\chi(A)}(\alpha)$ for some $\alpha = [U/\sim_A]$, so that $W =_o U$ by the Lemma. Thus,

$$W <_o \chi(A) \implies W \leq_c A, \tag{7.33}$$

since (the field of) U is a subset of A and similarities are injections. Taking the negation of both sides,

$$W \not\leq_c A \implies \neg[W <_o \chi(A)] \implies \chi(A) \leq_o W. \quad \neg$$

Of course, we would like to prove that $A <_c \chi(A)$ instead of the timid $\chi(A) \not\leq_c A$, and this is certainly true, but its proof depends on the Axiom of Choice. Wait for a bit until we finally bring the Deus ex Machina onto the stage.

The annoying details of this proof are forced on us by the fact that the restriction $\lesssim_A$ of the definite condition $\leq_o$ to $WO(A)$ is not a wellordering, for the trivial reason that it is not antisymmetric: there may exist distinct, similar $U, V \in WO(A)$, in fact they always do, if A has more than one element. This is why we were forced to take $h(A)$ as a set of equivalence classes rather than simply set $h(A) = WO(A)$. Technically, $\lesssim_A$ is a *prewellordering* (ugh!) and it is worth recasting the argument in a different form, after introducing this notion. See Problems **x7.15** - **x7.18**.

The Hartogs operation can be used to construct a general *supremum* operation for families of well ordered sets (Problem ***x7.24**), and it has many interesting properties. We use it next to extend the Continuous Least Fixed Point Theorem **6.21** to discontinuous mappings. Let us first put down, for the record, the Fixed Point Theorem for expansive mappings, which we have already discussed.

7.35. Fixed Point Theorem (Zermelo).² *Every expansive mapping $\pi : P \rightarrow P$ on an inductive poset has at least one fixed point, i.e. some $x^* \in P$ satisfies the equation*

$$x^* = \pi(x^*).$$

Proof. The argument given in **7.26** needs only some well ordered set U which cannot be injected into P , and $U = \chi(P)$ does it. $\dashv$

7.36. Least Fixed Point Theorem. *Every monotone mapping $\pi : P \rightarrow P$ on an inductive poset has exactly one strongly least fixed point x^* which is characterized by the two properties:*

$$\pi(x^*) = x^*, \tag{7.34}$$

$$(\forall y \in P)[\pi(y) \leq y \implies x^* \leq y]. \tag{7.35}$$

Proof. A careful examination of the proof of the Iteration Lemma **7.25** and the proof of the Fixed Point Theorem **7.26** reveals that exactly the same construction of the fixed point for an expansive mapping works and yields the least fixed point of a monotone mapping. However, it is not necessary to do this, as the Least Fixed Point Theorem is an easy consequence of the Fixed Point Theorem. The basic idea is to observe that the given monotone mapping π is necessarily expansive on some inductive *sub-poset* of P .

Let

$$Q = \{x \in P \mid x \leq \pi(x) \text{ \& } (\forall y)[\pi(y) \leq y \implies x \leq y]\}$$

²Zermelo did not formulate the Fixed Point Theorem in this generality, which is why it and many of its Corollaries have been attributed at various times to later mathematicians. But the famous “first proof” of the *Wellordering Theorem* which Zermelo gave in 1904 proves exactly this result, trivially restricted to the special case which interested him.

and observe first that the restriction

$$\leq_Q = \{(x, y) \mid x, y \in Q \text{ \& } x \leq_P y\}$$

of $\leq_P$ to Q is also a partial ordering—this is automatically true for the restriction of $\leq_P$ to any subset of P . (We skip the subscripts P and Q for the remaining of the argument.) In addition, $\pi[Q] \subseteq Q$, because

$$x \leq \pi(x) \implies \pi(x) \leq \pi(\pi(x)),$$

and for every y ,

$$\pi(y) \leq y \text{ \& } x \leq y \implies \pi(x) \leq \pi(y) \leq y,$$

by the monotonicity of π . It follows that the restriction

$$\pi_Q = \{(x, \pi(x)) \mid x \in Q\}$$

of π to Q is a mapping on Q , it continues to be monotone (of course) and it is also expansive, because of the definition of Q . To apply the Fixed Point Theorem 7.35 to Q and π_Q we need the following.

Lemma. *The poset Q is inductive.*

Proof. It is enough to show that for every chain $S \subseteq Q$, the least upper bound $M = \sup S$ (which exists in P because P is inductive) is a member of Q , i.e. (by the definition), (1) $M \leq \pi(M)$, and (2) for every y , $\pi(y) \leq y \implies M \leq y$. For (1) we compute:

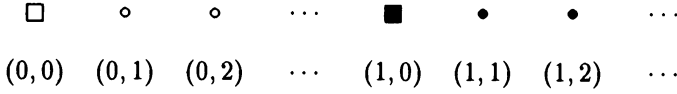
$$\begin{array}{lll} x \in S & \implies & x \leq M & \text{because } M \text{ is an upper bound of } S, \\ & \implies & \pi(x) \leq \pi(M) & \text{because } \pi \text{ is monotone,} \\ & \implies & x \leq \pi(x) \leq \pi(M) & \text{because } x \in S \subseteq Q, \end{array}$$

and therefore $\pi(M)$ is an upper bound of S and we have $M = \sup S \leq \pi(M)$. (2) follows from the observation that every y such that $\pi(y) \leq y$ is an upper bound of Q (from Q 's definition), and therefore an upper bound of the smaller set $S \subseteq Q$, so that $M = \sup S \leq y$.

By the Fixed Point Theorem 7.35 now, there exists some $x^* \in Q$, such that $\pi(x^*) = x^*$ and (7.35) holds simply because $x^* \in Q$.

The uniqueness of x^* is immediate from (7.34) and (7.35). ⊣

The full Least Fixed Point Theorem frees us from the necessity to check continuity in the applications of least fixed points to computer science, the fixpoint theory of programs. This is nice. It has, however, more significant, deeper applications to the general theory of sets, particularly in the study of *definability* in set theory as well as the construction of examples and counterexamples with specified properties. We will encounter several of these in the chapters which follow.

Figure 7.5. The sum $N +_o N$.

Problems

x7.1. Every linear ordering of a finite set is a wellordering. (See the related Problem *x6.6.)

7.37. The **sum** $P +_o Q$ of two posets P and Q is obtained by placing disjoint copies of P and Q side-by-side, every point of P preceding every point of Q . Formally, we set $P +_o Q = R$, where

$$\text{Field}(R) =_{\text{df}} (\{0\} \times \text{Field}(P)) \cup (\{1\} \times \text{Field}(Q)), \quad (7.36)$$

and for $(i, x), (j, y) \in \text{Field}(R)$,

$$(i, x) \leq_R (j, y) \iff_{\text{df}} i < j \vee [i = j = 0 \ \& \ x \leq_U y] \vee [i = j = 1 \ \& \ x \leq_V y]. \quad (7.37)$$

The idea is that P is similar with the set $\{0\} \times \text{Field}(P)$ partially ordered by its second elements, by the obvious similarity $(x \mapsto (0, x))$, and again $Q =_o \{1\} \times \text{Field}(Q)$.

x7.2. If $P =_o P'$ and $Q =_o Q'$, then $P +_o Q =_o P' +_o Q'$.

x7.3. For all posets P , $\text{Succ}(P) =_o P +_o [0, 1)$.

x7.4. For all posets P, Q, R ,

$$P +_o (Q +_o R) =_o (P +_o Q) +_o R.$$

x7.5. If U and V are well ordered sets, then so is their sum $U +_o V$.

x7.6. Prove that $[0, 1) +_o N =_o N \neq_o N +_o [0, 1)$, so that the addition operation on well ordered sets is not commutative.

7.38. The **product** $P \cdot_o Q$ of two posets is obtained by replacing each point of Q by a copy of P . Formally, we let $P \cdot_o Q = R$, where

$$\text{Field}(R) = \text{Field}(P) \times \text{Field}(Q), \quad (7.38)$$

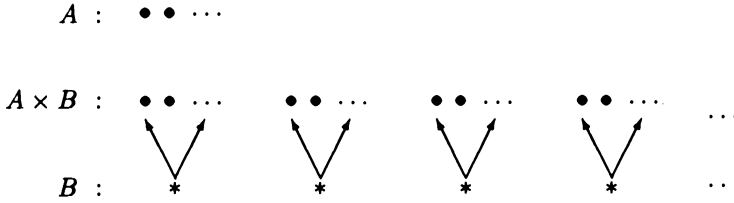


Figure 7.6. The product of two well ordered sets.

and $\leq_R$ is the *inverse lexicographic* ordering of pairs i.e. we compare the second members first: for $(x_1, y_1), (x_2, y_2) \in \text{Field}(R)$,

$$(x_1, y_1) \leq_R (x_2, y_2) \iff_{\text{df}} y_1 <_Q y_2 \vee [y_1 = y_2 \ \& \ x_1 \leq_P x_2]. \quad (7.39)$$

There is no special reason for ordering pairs by looking at their second members first, it is just that Cantor chose to do it this way and it has stuck.

x7.7. If $P =_o P'$ and $Q =_o Q'$, then $P \cdot_o Q =_o P' \cdot_o Q'$.

x7.8. Prove that $P \cdot_o [0, 2) =_o P +_o P$, but $[0, 2) \cdot_o N =_o N \neq_o N \cdot_o [0, 2)$, so that multiplication of well ordered sets is not commutative.

x7.9. For all posets P, Q, R ,

$$P \cdot_o (Q \cdot_o R) =_o (P \cdot_o Q) \cdot_o R.$$

x7.10. The product of two well ordered sets is well ordered.

x7.11. For each well ordered set U , there exists a unique function *Parity* : $U \rightarrow N$, such that *Parity*(y) = 0 if $y = 0$ or y is a limit point, and at successor points,

$$\text{Parity}(S(x)) = 1 - \text{Parity}(x).$$

x7.12. Every point y in a well ordered set U can be expressed uniquely in the form

$$y = S^n(x), \quad (7.40)$$

where (1) x is either the minimum 0 or a limit point, (2) n is a natural number and (3) the function $(i, x) \mapsto S^i(x)$ is defined by the recursion

$$S^0(x) = x, \quad S^{i+1}(x) = S(S^i(x)).$$

x7.13. For any two well ordered sets U, V , there exists at most one initial similarity $\pi : U \rightarrow \pi[U] \subseteq V$.

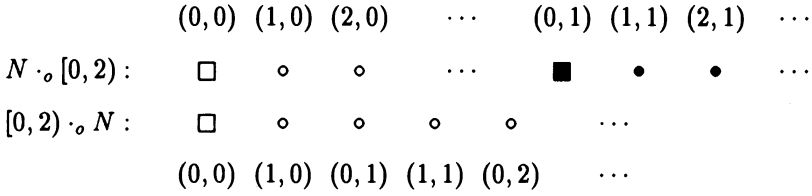


Figure 7.7. Multiplication of well ordered sets is not commutative.

x7.14. For all well ordered sets U, V, W ,

$$\begin{aligned} U <_o V \ \& \ V \leq_o W &\implies U <_o W, \\ U \leq_o V \ \& \ V <_o W &\implies U <_o W. \end{aligned}$$

7.39. Definition. A **prewellordering** on a set A is any relation $\lesssim \subseteq A \times A$ which is reflexive, transitive, connected (total) and grounded. “Connected” means that any two points in A are comparable,

$$(\forall x, y \in A)[x \lesssim y \vee y \lesssim x],$$

and “grounded” means that every non-empty $X \subseteq A$ has a $\lesssim$ -least member,

$$(\forall X \subseteq A, X \neq \emptyset)(\exists x \in X)(\forall y \in X)[x \lesssim y].$$

A prewellordering would be a wellordering, if only it were antisymmetric.

x7.15. For each set A , consider the set

$$B = \{X \subseteq A \mid X \text{ is finite}\}$$

of all finite subsets of A and set on B

$$X \lesssim_B Y \iff_{\text{df}} X \leq_c Y.$$

Prove that $\lesssim_B$ is a prewellordering.

x7.16. A relation $\lesssim \subseteq A \times A$ is a prewellordering if and only if there exists a well ordered set $U = (Field(U), \leq_U)$ and a surjection $\pi : A \twoheadrightarrow Field(U)$ such that

$$x \lesssim y \iff \pi(x) \leq_U \pi(y) \quad (x, y \in A).$$

x7.17. For each set A , the relation

$$U \lesssim_A V \iff_{\text{df}} U, V \in WO(A) \ \& \ U \leq_o V$$

is a prewellordering of $WO(A)$.

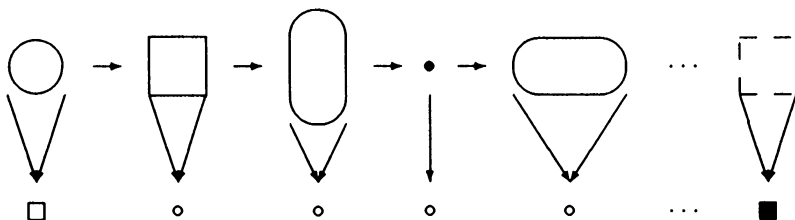


Figure 7.8. Portrait of a prewellordering.

x7.18. Rework the proof of Hartogs' Theorem by applying the preceding two problems.

x7.19. For every set A , there exists a well ordered set V such that there exists no surjection $\pi : A \rightarrow V$.

x7.20. Prove that $A \leq_c B \implies \chi(A) \leq_o \chi(B)$.

x7.21. Prove that $\chi([0, n)) =_o [0, n + 1)$.

x7.22. If W is a well ordered set and $W \leq_c A$, then $W <_o \chi(A)$.

x7.23. For each set A and each well ordered set U ,

$$U <_o \chi(A) \iff \text{Field}(U) \leq_c A.$$

***x7.24.** The operation $\chi(A)$ is definite, we gave an explicit definition of the field $h(A)$ and the wellordering $\leq_{\chi(A)}$ of $\chi(A)$ from A . Define a similar operation $\sup(\mathcal{E})$, such that if $\mathcal{E}$ is a non-empty family of well ordered sets, then:

1. $\sup(\mathcal{E})$ is a well ordered set.
2. $U \in \mathcal{E} \implies U \leq_o \sup(\mathcal{E})$.
3. If W is a well ordered set and for each $U \in \mathcal{E}$, $U \leq_o W$, then $\sup(\mathcal{E}) \leq_o W$.

***x7.25.** Let $\leq$ be a linear ordering of a set A and define on the poset $\mathcal{P}(A)$ the mapping

$$\pi(X) =_{\text{df}} \{y \in A \mid (\forall x < y)[x \in X]\}.$$

Verify that $\pi : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$ is monotone and give an example where it is not countably continuous. Prove that if A_w is the least fixed point of π , then

$$x \in A_w \iff \{(s, t) \in A \times A \mid s \leq t < x\} \text{ is a wellordering.}$$

There are situations where it is easier to use *the proof* of the Fixed Point Theorem 7.35 rather than its statement.

***x7.26. Detailed Fixed Point Theorem.** For each expansive or monotone mapping $\pi : P \rightarrow P$ on an inductive poset P , there exists a subset $D \subseteq P$ with the following properties:

1. D is a well ordered chain in P .
2. Every member of D is determined from its predecessors by the formula

$$x = \pi(\sup \{y \in D \mid y < x\}).$$

3. No point in D is a fixed point of π .
4. The point $\pi(\sup D)$ is a fixed point of π ,

$$\pi(\pi(\sup D)) = \pi(\sup D).$$

5. If π is monotone, then $\pi(\sup D)$ is the least fixed point of π .

Prove also that these conditions determine D uniquely.

***x7.27.** Suppose P and Q are inductive posets and $\pi : P \times Q \rightarrow P \times Q$ is a monotone mapping on the product and define the mapping $\rho : Q \rightarrow P$ by appealing to Problem x6.4 and the Least Fixed Point Theorem 7.36,

$$\begin{aligned} \rho(y) &= (\mu x \in P)[\pi(x, y) = x] \\ &= \text{the least fixed point of } \pi(x, y) = x. \end{aligned} \quad (7.41)$$

Prove that ρ is a monotone mapping, and if π is countably continuous, then so is ρ .

***x7.28. Bekiĉ-Scott Rule.** Suppose P_1, P_2 are inductive posets, and

$$\pi_1 : P_1 \times P_2 \rightarrow P_1, \quad \pi_2 : P_1 \times P_2 \rightarrow P_2$$

are monotone mappings. Using the μ -notation for least fixed points of (7.41), let

$$\rho(x_2) = (\mu x_1 \in P_1)[\pi_1(x_1, x_2) = x_1],$$

let

$$\bar{x}_2 = (\mu x_2 \in P_2)[\pi_2(\rho(x_2), x_2) = x_2]$$

be the least fixed point of the mapping $x_2 \mapsto \pi_2(\rho(x_2), x_2)$ (which is monotone by *x7.27) and finally let

$$(x_1^*, x_2^*) = (\mu(x_1, x_2) \in P_1 \times P_2)[(\pi_1(x_1, x_2), \pi_2(x_1, x_2)) = (x_1, x_2)]$$

be the least fixed point in the product poset. Prove that

$$x_2^* = \bar{x}_2.$$

The problem insures that we can compute simultaneous least fixed points by iterating the least fixed point operation ($\mu x \in P$) on one inductive poset at a time.

Chapter 8

CHOICES

8.1. The Axiom of Choice, AC. For any two sets A , B and any binary relation $P \subseteq (A \times B)$,

$$(\forall x \in A)(\exists y \in B)[xPy] \implies (\exists f : A \rightarrow B)(\forall x \in A)[xPf(x)]. \quad (8.1)$$

This is the last and most controversial axiom of Zermelo. To understand how such an axiom might be needed, consider the classical example of Russell, where A is a set of pairs of shoes, $B = \bigcup A$ and

$$xPy \iff y \in x.$$

The function

$$f(x) =_{\text{df}} \text{the left shoe of } x \quad (x \in A)$$

obviously selects a shoe from each pair, in symbols $(\forall x \in A)[xPf(x)]$. If, however, A is a set of pairs of *socks*, then we cannot *define* a function $f : A \rightarrow \bigcup A$ which selects one sock $f(x) \in x$ from each pair, because (as we stipulate for the example), a pair of socks comprises precisely two perfectly identical objects. We can still prove that a selector function f exists when A is finite, by induction on the number of elements in A (Problem **x8.1**). But in mathematics we can imagine infinite sets of pairs of socks, and in that case we need something like the Axiom of Choice to guarantee the existence of such a function.

Less amusing but more significant for mathematics is the proof of the basic theorem **2.10**, where we consider a sequence $A_0, A_1, \dots$ of countable sets and we begin with the phrase

It is enough to prove the theorem in the special case where none of the A_n is empty, in which case we can find for each A_n an enumeration $\pi_n : N \rightarrow A_n$.

Perhaps for each n “we can find” (i.e. “there exists”) some enumeration π of A_n , but the rest of the proof needs a *function* ($n \mapsto \pi_n$) which associates a specific enumeration π_n with each n : which of the axioms **(I)** - **(VI)** can

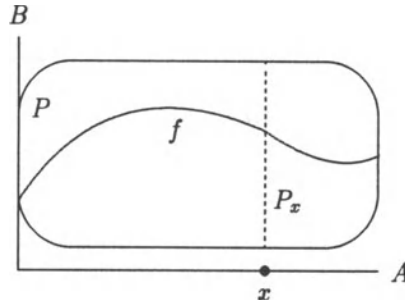


Figure 8.1. A selector for $P \subseteq A \times B$.

be used to prove the existence of such a function? Here $A = N$, $B = (N \rightarrow \bigcup_{n=0}^{\infty} A_n)$ and

$$nP\pi \iff \pi : N \rightarrow A_n,$$

so that $(\forall n \in N)(\exists \pi \in B)[nP\pi]$ from the hypothesis that each A_n is non-empty and countable, and the Axiom of Choice guarantees precisely that there exists a function $f : N \rightarrow B$ such that for each $n \in N$, the value $f(n) = \pi_n$ satisfies $nP\pi_n$, i.e. it enumerates A_n . Such “silent” appeals to the Axiom of Choice are very common in mathematics and especially in analysis, where the classical theory of limits and continuous functions cannot be developed in a satisfactory way without choices.

If we picture $P \subseteq A \times B$ as a subset of the product space, then the hypothesis $(\forall x \in A)(\exists y \in B)[xPy]$ means that the *fiber* or *section*

$$P_x =_{\text{df}} \{y \in B \mid xPy\} \quad (8.2)$$

above each $x \in A$ is non-empty; the Axiom of Choice guarantees the existence of a *selector* for P , a function $f : A \rightarrow B$ which assigns to each $x \in A$ exactly one point in the fiber above it. There are two other, simple reformulations of the axiom which express in different ways the process of “collecting into a whole” any number of unrestricted, non-conflicting choices.

8.2. Definition. A set S is a **choice set** for a family of sets $\mathcal{E}$ if (1) $S \subseteq \bigcup \mathcal{E}$, and (2) for every $X \in \mathcal{E}$, the intersection $S \cap X$ is a singleton. A choice set S selects from each $X \in \mathcal{E}$ the unique member of the intersection $S \cap X$.

8.3. Exercise. If $\emptyset \in \mathcal{E}$, then $\mathcal{E}$ does not admit a choice set. Also, if $a \neq b$, then the family $\mathcal{E} = \{\{a\}, \{a, b\}, \{b\}\}$ does not admit a choice set.

8.4. Theorem. The Axiom of Choice is equivalent to the following proposition: every family $\mathcal{E}$ of non-empty and pairwise disjoint sets admits a choice set. This is the version of **AC** postulated by Zermelo.

Proof. Assume first the Axiom of Choice and let $U = \bigcup \mathcal{E}$ be the union of the given family of pairwise disjoint, non-empty sets, which means that

$$(\forall X \in \mathcal{E})(\exists x \in U)[x \in X].$$

The Axiom of Choice guarantees that there exists a function $f : \mathcal{E} \rightarrow U$, such that

$$(\forall X \in \mathcal{E})[f(X) \in X];$$

we set $S = f[\mathcal{E}] = \{f(X) \mid X \in \mathcal{E}\}$ and the fact that the members of $\mathcal{E}$ are pairwise disjoint implies easily that S intersects every member of $\mathcal{E}$ in a singleton. For the converse, assume

$$(\forall x \in A)(\exists y \in B)[x P y],$$

set for each $x \in A$

$$U_x = \{(t, y) \mid t P y \text{ \& } t = x\}$$

and let

$$\mathcal{E} = \{U_x \mid x \in A\}.$$

Each member of $\mathcal{E}$ is non-empty by the hypothesis and it is determined by the constant, first member of the pairs in it, so any two members of $\mathcal{E}$ are disjoint. If S is a choice set for this $\mathcal{E}$, then the function

$$f(x) = \text{the unique } y \text{ such that } (x, y) \in S$$

easily satisfies the conclusion of the Axiom of Choice. $\dashv$

8.5. Definition. A choice function for a set A is any partial function $\varepsilon : \mathcal{P}(A) \rightarrow A$, such that

$$\emptyset \neq X \subseteq A \implies \varepsilon(X) \downarrow \text{ \& } \varepsilon(X) \in X.$$

8.6. Lemma. The Axiom of Choice is equivalent to the assertion that every set admits a choice function.

Proof. For every A , obviously

$$(\forall X \in \mathcal{P}(A) \setminus \{\emptyset\})(\exists y \in A)[y \in X],$$

and directly from the Axiom of Choice, there must exist some function $\varepsilon : \mathcal{P}(A) \setminus \{\emptyset\} \rightarrow A$ such that

$$(\forall X \in \mathcal{P}(A) \setminus \{\emptyset\})[\varepsilon(X) \in X].$$

The converse is easy enough to leave for an exercise. $\dashv$

8.7. Exercise. If every set admits a choice function, then the Axiom of Choice is true.

8.8. But is it true? (1) Naively understood, the Axiom of Choice asserts that if each of a set of non-conflicting choices is possible, then they can all be made independently and their results collected into a completed whole, a set. By this understanding it is quite obvious, it can be justified by the natural interpretation we would give to the Powerset Axiom: when we grant sethood to the class $\{X \mid X \subseteq A\}$ of all subsets of A , we truly mean *all* subsets of A , including those for which the membership criterion is not determined by some explicit law but by free choice, by chance if you will.

The Axiom of Choice is different in form from the earlier “constructive” axioms (II) - (VI), because it *postulates directly* the existence of a set for which it does not supply a definition. Each of (II) - (VI) grants sethood to a specific, explicitly defined collection of objects, it legitimizes a special case of the most appealing (if false) General Comprehension Principle 3.3. *The Axiom of Choice is the only Zermelo axiom other than Extensionality which is not a special case of the General Comprehension Principle.* This is misstated on occasion, to make the claim that the Axiom of Choice is the only one which *demand the existence* of objects for which it does not supply a definition, which is not true: the Extensionality and Powerset Axioms do the same, in a more fundamental if indirect manner.

Zermelo introduced the Axiom of Choice explicitly in 1904, in a brief paper in which he used it to prove that every set is well orderable. This was a long-standing conjecture, and Cantor had outlined a proof of it in a letter to Dedekind, then still unpublished. His proof, however (and the related proof of the Cardinal Comparability Hypothesis), depended on intuitions about sets which were not sufficiently explained. In contrast to this, Zermelo made it clear, from the start, that his own detailed proof depended on the Axiom of Choice, and he was immediately attacked for this by some of the leading mathematicians of the time, for introducing a questionable method to derive an implausible conclusion. Given the fact that choice principles were by no means new to mathematics and that they permeate Cantor’s earlier reasoning, it is fair to say that the shock was caused more by the realization of the power of the axiom than by its meaning.

In the next theorem we list some of the more famous propositions about sets which are equivalent to the Axiom of Choice. We have used the traditional names for these propositions, Lemma, Hypothesis, Theorem, which have been attached to them by the historical accident of when and how they were introduced in the mathematical literature.

8.9. Theorem. *The following propositions are all equivalent.*

(1) **Axiom of Choice.**

(2) **Maximal Chain Principle:** *every poset P has a chain $S \subseteq P$ which is maximal, in the sense that for every other chain S' , $S \subseteq S' \implies S = S'$.*

(3) **Zorn's Lemma:** *if every chain in a poset P has an upper bound, then P has at least one maximal element.*

(4) **Hypothesis of Cardinal Comparability:** *for any two sets A, B , either $A \leq_c B$ or $B \leq_c A$.*

(5) **Wellordering Theorem:** *every set is well orderable.*

Proof. We verify, round-robin style, that each of these propositions implies the next and finally (5) $\implies$ (1).

(1) $\implies$ (2). The poset $\text{Chains}(P)$ is inductive, 6.14. If P has no maximal chain, then every chain has a proper extension; the Axiom of Choice gives us a function $\pi : \text{Chains}(P) \rightarrow \text{Chains}(P)$ such that for all S , $S \subsetneq \pi(S)$; and π is an expansive mapping with no fixed point in $\text{Chains}(P)$, contradicting the Fixed Point Theorem 7.35.

(2) $\implies$ (3). By (2), P has a maximal chain S , and by hypothesis, S has an upper bound M which is maximal in P ; because if $M < y$, then $S \cup \{y\}$ is a chain properly extending S , which does not exist.

(3) $\implies$ (4) Every chain S in the poset $(A \curvearrowright B)$ of all partial injections on A has an upper bound, namely, its union $\bigcup S$. Thus, by (3), there exists a maximal partial injection $f : A \curvearrowright B$. If

$$a \in A \setminus \text{Domain}(f), \quad b \in B \setminus f[\text{Domain}(f)],$$

then $f \cup \{(a, b)\}$ is (easily) a partial injection which properly extends f ; thus either $f[A] = B$ and f is total and witnesses $A \leq_c B$, or $f[\text{Domain}(f)] = B$ and the inverse partial injection $f^{-1} : B \curvearrowright A$ is total and witnesses that $B \leq_c A$.

(4) $\implies$ (5). Given A , let $h(A)$ be the Hartogs set associated with A , which is well ordered by $\leq_{\chi(A)}$. By Hartogs' Theorem 7.34, $A \not\leq_c h(A)$, so (4) guarantees the existence of an injection $f : A \hookrightarrow h(A)$ and we can define on A the relation

$$x \leq y \iff_{\text{df}} f(x) \leq_{\chi(A)} f(y).$$

This is easily a wellordering.

(5) $\implies$ (1). If $\leq$ is a wellordering of A , then the partial function

$$\varepsilon(X) = \text{the } \leq\text{-least member of } X$$

is a choice function for A . ⊣

8.10. But is it true? (2) The meaning of this theorem is that if we accept the basic, constructive first six axioms of Zermelo, then the Axiom of Choice, the Maximal Chain Principle, Zorn's Lemma, the Hypothesis of

Cardinal Comparability and the Wellordering Theorem express in five different ways the same set theoretic principle. No doubt, the Axiom of Choice is the most direct and intuitive formulation of this principle, the one which makes it most obvious that it is true. The Maximal Chain Principle and Zorn's Lemma are technical, abstract and hard to understand, but they are valuable for their applications, especially in analysis, algebra and topology. We will not do much with these propositions in the main body of these Notes and we have included them mostly because the mathematicians who use set theory rather than do it swear by them. The Cardinal Comparability Hypothesis is certainly easy to understand and plausible, but few would propose it as an axiom, it has the feel of a proposition which ought to be proved. Finally, the Wellordering Theorem is crystal clear in its meaning and it gives a mechanism for making choices which "explains" in some way the Axiom of Choice, but far from being obvious, it raises a flag of caution. For example, what does a wellordering of the powerset of the natural numbers $\mathcal{P}(N)$ look like? Without some thought it is not even obvious that $\mathcal{P}(N)$ admits linear orderings (see Problem *x8.9). It is quite difficult to imagine the structure of the beast, and this naturally casts doubt on the truth of the axiom which implies its existence. It is hardly surprising that the commotion about the Axiom of Choice was caused by Zermelo's proof of the implication (1) $\implies$ (5), whose conclusion is still thought by many to be counterintuitive.

We now consider two easy corollaries of the Axiom of Choice which express simpler principles of choice.

8.11. Countable Principle of Choice, $\mathbf{AC}_N$. *For each set B and each binary relation $P \subseteq N \times B$ between natural numbers and members of B ,*

$$(\forall n \in N)(\exists y \in B)[nPy] \implies (\exists f : N \rightarrow B)(\forall n \in N)[nPf(n)].$$

8.12. (VI) Axiom of Dependent Choices, $\mathbf{DC}$. *For each set A and each relation $P \subseteq A \times A$,*

$$\begin{aligned} a \in A \ \& \ (\forall x \in A)(\exists y \in A)[xPy] \\ \implies & (\exists f : N \rightarrow A)[f(0) = a \ \& \ (\forall n \in N)[f(n)Pf(n+1)]]. \end{aligned}$$

In contrast to the full Axiom of Choice which demands the existence of choice functions $f : A \rightarrow B$ for arbitrary A, B , the Countable Principle of Choice $\mathbf{AC}_N$ justifies only a sequence of independent choices from an arbitrary set B which successively satisfy the conditions

$$0Pf(0), \ 1Pf(1), \ 2Pf(2), \ \dots$$

The Axiom of Dependent Choices also justifies only a sequence of choices, where, however, each of them may depend on the previous one, since they

must now satisfy the conditions

$$f(0)Pf(1), f(1)Pf(2), f(2)Pf(3), \dots$$

It is easily equivalent to the following, seemingly stronger principle which allows each choice to depend on all the preceding ones.

8.13. Proposition. *The Axiom of Dependent Choices is equivalent to the following proposition: for every set A and every relation $P \subseteq A^* \times A$ between strings from A and members of A ,*

$$(\forall u \in A^*)(\exists x \in A)[uPx] \implies (\exists f : N \rightarrow A)(\forall n)[\bar{f}(n)Pf(n+1)].$$

Proof. The implication from this version of **DC** to the “official” one is easy and we leave it for an exercise. Assuming now **DC** and the hypothesis of the seemingly stronger version, define on A^* the relation

$$uQv \iff_{\text{df}} (\exists x \in A)[v = u * \langle x \rangle \ \& \ uPx];$$

we obviously have $(\forall u \in A^*)(\exists v \in A^*)[uQv]$, **DC** gives us a function $g : N \rightarrow A^*$ such that $g(0) = \emptyset$ and $(\forall n)[g(n)Qg(n+1)]$, and the function we need is (easily) $f = \bigcup g$. $\dashv$

8.14. Theorem. (1) *The Axiom of Choice implies the Axiom of Dependent Choices.*

(2) *The Axiom of Dependent Choices implies the Countable Principle of Choice.*

Proof. (1) Let $\varepsilon : \mathcal{P}(A) \setminus \{\emptyset\} \rightarrow A$ be a choice function for A and assume the hypothesis of **DC**. The function $f : N \rightarrow A$ that we need for the conclusion is defined by the recursion

$$\begin{aligned} f(n) &= a, \\ f(n+1) &= \varepsilon(\{y \in A \mid f(n)Py\}). \end{aligned}$$

(2) Assume the hypothesis of the Countable Principle of Choice, let $A = N \times B$, let $a = (0, b)$ where $b \in B$ is any point satisfying $0Pb$ and define on A the relation

$$(n, x)Q(m, y) \iff_{\text{df}} m = n + 1 \ \& \ mPy.$$

The function $f : N \rightarrow N \times B$ supplied by **DC** for this a and Q takes pairs as values, so $f(n) = (g(n), h(n))$, $g(0) = 0$, $h(0) = b$ for suitable functions g, h , and for every n , $g(n+1) = g(n) + 1$, $g(n+1)Ph(n+1)$. It follows that for every n , $g(n) = n$ and $nPh(n)$, as required by the conclusion of the Countable Principle of Choice. $\dashv$

We need a definition to formulate the most basic version of the Axiom of Dependent Choices.

8.15. Definition. A graph $(G, \rightarrow_G)$ is **grounded** or **well founded** if every non-empty subset of G has a minimal member:

$$\emptyset \subsetneq S \subseteq G \implies (\exists m \in S)(\forall x \in S)[\neg m \rightarrow_G x]. \quad (8.3)$$

A poset $(P, \leq)$ is **grounded** if the associated “inverse strict graph” $(P, >)$ is grounded, which means that for every S ,

$$\emptyset \subsetneq S \subseteq P \implies (\exists m \in S)(\forall x \in S)[x \leq m \implies x = m]. \quad (8.4)$$

8.16. Exercise. Assume **DC** and prove that a linear ordering $(P, \leq)$ is grounded if and only if it is a wellordering.

8.17. Proposition. The Axiom of Dependent Choices is equivalent to the following proposition: a graph G is grounded if and only if it has no **infinite, descending chains**, i.e. there exists no function $f : N \rightarrow G$ such that for all n , $f(n) \rightarrow_G f(n+1)$,

$$f(0) \rightarrow_G f(1) \rightarrow_G f(2) \rightarrow_G \dots$$

Proof. First assume **DC**. If $f : N \rightarrow G$ is an infinite, descending chain, then the set $\{f(n) \mid n \in N\}$ has no minimal element, so G is not grounded. Conversely, if G has a non-empty subset A with no minimal element, then $(\forall x \in A)(\exists y \in A)[x \rightarrow_G y]$, and then **DC** gives us an infinite descending chain.

Assume now that every graph which has no infinite descending chains is grounded and the hypothesis

$$(\forall x \in A)(\exists y \in A)[xPy]$$

of **DC** holds, and consider the graph $(A, \rightarrow_A)$ where

$$x \rightarrow_A y \iff_{\text{df}} xPy \quad (x, y \in A).$$

The conclusion of **DC** is exactly the statement that $(A, \rightarrow_A)$ has an infinite descending chain, so if it fails, there must exist some minimal $s \in A$; this means precisely that $(\forall y \in A)\neg[sPy]$, which contradicts the hypothesis of **DC**. $\neg$

Grounded graphs have many of the properties of well ordered sets, in particular, we can prove propositions by induction and define functions by recursion over them, Problems **x8.10** and ***x8.11** and Theorem **11.5**. The easy direction of this result makes **DC** particularly useful in studying them, as it is often simpler to verify that a given graph G has no infinite descending chains than to prove directly that G is grounded.

8.18. But is it true? (3) We have remarked that before it was formulated precisely by Zermelo, the Axiom of Choice had been used many times “silently” in classical mathematics, and in particular in analysis. *These classical applications, however, can all be justified on the basis of the Axiom of Dependent Choices*, in fact most of them need only the weaker Countable Principle of Choice. This will become clear in Chapter 10 and Appendix A. Zermelo assumed the full Axiom of Choice because it is a natural hypothesis in the context of Cantor’s set theory; because it is needed in the proofs of the Wellordering Theorem and the Cardinal Comparability Hypothesis; and because it is indispensable for the development of cardinal arithmetic. This difference between the choice principles needed for classical mathematics and those required by Cantor’s new theory of sets explains in part the strident reaction to the axioms of Zermelo by the distinguished analysts of his time (including the great Borel), who had used choice principles routinely in their work—and continued using them, as they denounced general set theory and called it an illusion: in the context of 19th century classical analysis, the Axiom of Dependent Choices is natural and necessary, while the full Axiom of Choice is unnecessary and even has some counterintuitive consequences, including certainly the Wellordering Theorem.

We should also mention here that even in general set theory where the full Axiom of Choice is routinely accepted as obvious, many of the basic theorems do not need it, and in particular *all the results of Chapter 3 can be based axiomatically on the Axiom of Dependent Choices*. Notice also that we proved all the basic facts about well ordered sets in the preceding chapter with no appeal to choice principles whatsoever. For this reason, we will deviate technically from Zermelo and we will put in our basic system the Axiom of Dependent Choices instead of the full Axiom of Choice. “Technically,” because we take the position that there is no doubt about the truth of the Axiom of Choice and we will never hesitate to appeal to it when it is needed, we will simply include it (discreetly) among the hypotheses.

8.19. Axiomatics: the theories ZDC, ZAC. *The axiomatic system ZDC comprises the constructive axioms (I) - (VI) of Chapter 3 and the Axiom (VII) of Dependent Choices 8.12. The classical system ZAC of Zermelo includes, in addition, the full Axiom of Choice, AC, 8.1. Symbolically*

$$\begin{aligned}\mathbf{ZDC} &= (\mathbf{I}) - (\mathbf{VI}) + \mathbf{DC} = (\mathbf{I}) - (\mathbf{VII}), \\ \mathbf{ZAC} &= (\mathbf{I}) - (\mathbf{VI}) + \mathbf{AC} = \mathbf{ZDC} + \mathbf{AC}.\end{aligned}$$

From now on and until Chapter 11, we will use in proofs the axioms of **ZDC** without explicit mention. When the Axiom of Choice is required, we will make a note of the fact by annotating the relevant proposition with the mark **(AC)**. In Chapter 11 we will complete our axiomatization by adding to **ZDC** the *Axiom of Replacement*.

8.20. Consistency and independence results. Could we settle the controversy about the Axiom of Choice by simply *proving* or *refuting* **AC** from the constructive axioms **(I)** - **(VI)**? Neither possibility seems likely. On the one hand, **AC** is probably true, as are axioms **(I)** - **(VI)**, and we cannot refute a true statement on the basis of true assumptions. On the other hand, **AC** appears to be a genuinely new set theoretic principle, and we cannot expect to prove it from the other ones, by logic alone. As a matter of fact, it can be shown rigorously that the Axiom of Choice can neither be proved nor refuted from axioms **(I)** - **(VI)**.

The most direct way to show that a certain proposition ϕ cannot be proved in a certain axiomatic system **T** is to produce a **model** of **T**, in which ϕ is false. Consider the classical problem about plane Euclidean geometry, whether the *Parallel Axiom*¹ can be deduced from the others. To show that it cannot, we declare that by “plane” we will mean the two-dimensional sphere, the surface of the unit ball, and by “line” we will mean any great circle on the sphere. The remaining primitive notions of plane Euclidean geometry can be defined naturally in this interpretation, and it is not hard to verify that the basic, simple axioms of Euclid are true with these definitions; thus, we have a model of plane geometry in which the Parallel Axiom fails, simply because any two great circles intersect. We conclude that the Parallel Axiom cannot be proved from the others “by logic alone,” because then it would be true in every interpretation which makes the other axioms true, and we have found one where it is false.

To define a model for an axiomatic theory, in general, one needs to specify a domain of objects and interpret on it the primitives of the theory, so that the axioms are true. For a theory about sets, this means we must define *sethood* and *membership* on some domain, and we must also identify which conditions and operations on the domain will be considered *definite*. Models for **ZDC** and **ZAC** do not come cheap, the theories are too strong. We will study some very special models (“universes”) in Appendix B, but the most interesting constructions require delicate methods from *mathematical logic* which are outside the scope of these Notes: we will just state and discuss some of the many famous consistency and independence results of the subject as they become relevant in what follows.

We have assumed at the outset, in **3.6**, that our theory has a model, the standard universe of objects $\mathcal{W}$, in which axioms **(I)** - **(VI)** (at least) are true. This assumption is natural and even necessary if our lives as set theorists are to have any meaning, but it is not included among the axioms of **ZDC**, **ZAC** or any of the other theories we will consider.² It is almost

¹The Parallel (fifth) Axiom of Euclid is equivalent to the assertion that given a line L and a point P not on it, there exists exactly one line L' through P and having no points in common with L .

²In fact it is not possible to assume such an axiom: adding the existence

never needed, except when we assert the existence of models of various theories: to construct those, we have to start with something, and that is always the assumed, standard model of our theory.

8.21. Proviso for model existence assertions. *Without further mention, all claims in these Notes of existence of models, consistency of theories and independence of propositions are based on the existence of a model which satisfies axioms (I) - (VI) and (VIII), the Axiom of Replacement which we will introduce in Chapter 11.*

8.22. The consistency of the Axiom of Choice (Gödel, 1939). *Zermelo's theory ZAC with the full Axiom of Choice has a model and hence (I) - (VI) do not refute AC, or AC is consistent with (I) - (VI).* Gödel's famous model L of *constructible sets* has many more canonical properties and it establishes the consistency of AC with theories much stronger than (I) - (VI). We will come back to it on several occasions.

8.23. The independence of the Axiom of Choice (Fraenkel-Mostowski, 1939, Cohen, 1963). *Each of the theories*

$$(I) - (VI) + \neg AC_N, (I) - (VI) + AC_N + \neg DC, ZDC + \neg AC$$

has a model. This means that we cannot prove AC_N from the constructive axioms (I) - (VI), we cannot prove DC from the constructive axioms and AC_N , and we cannot prove AC in ZDC: each of these three choice principles is stronger than the preceding ones. The early model constructions of Fraenkel and Mostowski either contained atoms or had some other, technical defects which limited the possibility of generalizing them. Cohen constructed his models by his famous *forcing method*, which he (and others) also used to establish many more unprovability results. We will refer to it several times in the remainder of these Notes.

Problems

Let us call two propositions ϕ and ψ **constructively equivalent** if their equivalence $\phi \iff \psi$ can be established on the basis of the constructive axioms (I) - (VI), i.e. without appealing to any choice principle whatsoever.

of a model of ZDC to the axioms of ZDC creates a new and stronger theory ZDC' and the further problem whether ZDC' has a model. In the best and most famous result of Mathematical Logic, Gödel proved (rigorously) that this conundrum cannot be avoided, *there exists no axiomatic theory (consistent and worth studying) which includes among its axioms or theorems the assertion that it possesses a model.*

x8.1. Prove the Axiom of Choice (8.1) for finite A .

Combined with Problem **x5.24**, the next problem gives a formulation of the Countable Principle of Choice $\mathbf{AC}_N$ directly in terms of the membership relation, with no reference to N or the concept of “function.”

x8.2. The Countable Principle of Choice $\mathbf{AC}_N$ is constructively equivalent to the following proposition: every countable, infinite family $\mathcal{E}$ of non-empty and pairwise disjoint sets admits a choice set.

x8.3. The Axiom of Choice is constructively equivalent to the following proposition: for every $A \neq \emptyset$ and every $f : A \rightarrow B$, there exists some $g : B \rightarrow A$ such that for all $x \in A$, $f(g(f(x))) = f(x)$.

x8.4. The Axiom of Choice is constructively equivalent to the following proposition: for each I and each indexed family of sets $(i \mapsto A_i)$ on I ,

$$(\forall i \in I)[A_i \neq \emptyset] \implies \prod_{i \in I} A_i \neq \emptyset.$$

The Countable Principle of Choice is constructively equivalent to the proposition: for every sequence of sets $(n \mapsto A_n)$, $n \in N$,

$$(\forall n \in N)[A_n \neq \emptyset] \implies \prod_{n \in N} A_n \neq \emptyset.$$

In the next four problems we establish that the Axiom of Dependent Choices is equivalent to several seemingly weaker principles of choice.

***x8.5.** The Axiom of Dependent Choices is constructively equivalent to the following proposition: for every non-empty A and every relation $P \subseteq A \times A$,

$$\begin{aligned} (\forall x \in A)(\exists y \in A)[xPy] \\ \implies (\exists B \subseteq A)[B \neq \emptyset \ \& \ (\exists f : B \rightarrow B)(\forall x \in B)[xPf(x)]] \end{aligned}$$

***x8.6.** The Axiom of Dependent Choices is constructively equivalent to the following proposition: for every relation $P \subseteq A \times A$ and $a \in A$,

$$\begin{aligned} (\forall x \in A)(\exists y \in A)[xPy] \\ \implies (\exists B \subseteq A)[a \in B \ \& \ (\exists f : B \rightarrow B)(\forall x \in B)[xPf(x)]] \end{aligned}$$

***x8.7.** The Axiom of Dependent Choices is constructively equivalent to the following proposition: a poset P is grounded if and only if it has no **infinite, descending chains**, i.e. if for every $f : N \rightarrow P$,

$$(\forall n \in N)[f(n+1) \leq f(n)] \implies (\exists n)[f(n+1) = f(n)].$$

It is also possible to formulate the Axiom of Dependent Choices directly in terms of the membership relation, but not in a very pretty manner.

***x8.8.** Prove that the following proposition is constructively equivalent to the Axiom of Dependent Choices: for every set A and every binary definite condition P ,

$$\begin{aligned} & [\emptyset \in A \ \& \ (\forall u, v \in A)[P(u, v) \implies (\exists! x)[v = u \cup \{x\}]] \\ & \ \& \ (\forall u \in A)(\exists v \in A)P(u, v)] \\ & \implies (\exists B \subseteq A)[\emptyset \in B \ \& \ (\forall u \in B)(\exists! v \in B)P(u, v)]. \end{aligned}$$

***x8.9.** Prove that the following, lexicographic ordering on $(N \rightarrow N)$ is indeed a linear ordering but not a wellordering:

$$f \leq g \iff_{\text{df}} f = g \vee (\exists n \in N)[(\forall i < n)[f(i) = g(i)] \ \& \ f(n) < g(n)].$$

Infer that $\mathcal{P}(N)$ admits a linear ordering.

x8.10. Grounded induction. For each grounded graph G and each unary definite condition P ,

$$(\forall y \in G)[(\forall x)(y \rightarrow_G x \implies P(x)) \implies P(y)] \implies (\forall y \in G)P(y).$$

***x8.11.** For every grounded graph G and every function

$$h : (G \multimap E) \rightarrow E,$$

there exists a unique (total) function $f : G \rightarrow E$ which satisfies the identity

$$f(x) = h(f \upharpoonright \{y \in G \mid x \rightarrow_G y\}) \quad (x \in G).$$

HINT. Rework the proof of Theorem 7.24, using functions

$$\sigma_t : \{x \in G \mid t \Rightarrow_G x\}$$

defined on “initial segments” of the transitive closure $\Rightarrow_G$ of $\rightarrow_G$.

Chapter 9

CHOICE'S CONSEQUENCES

We will begin this chapter with a few results about countability whose proofs illustrate the difference between $\mathbf{AC}_N$, $\mathbf{DC}$ and $\mathbf{AC}$, but our main task is to establish some important consequences of the full Axiom of Choice, including the basic laws of *cardinal arithmetic*. The telltale mark ($\mathbf{AC}$) will grace practically all the numbered propositions.

9.1. Theorem. *Every infinite set has a countable, infinite subset, so for every cardinal number κ , either $\kappa <_c \aleph_0$ or $\aleph_0 \leq_c \kappa$.*

Proof. If A is infinite, obviously

$$(\forall u \in A^*)(\exists y \in A)(\forall i < lh(u))[u(i) \neq y].$$

It follows from $\mathbf{DC}$ that there exists a sequence $f : N \rightarrow A$ such that

$$(\forall n)(\forall i < n)[f(i) \neq f(n)],$$

and the image $f[N]$ is a countable, infinite subset of A . The second assertion is trivial, taking cases on whether κ is finite or infinite. $\dashv$

The point of the second assertion of the theorem is that while the general property of Cardinal Comparability requires the full Axiom of Choice, the special (and significant) case of comparability with $\aleph_0$ is a theorem of $\mathbf{ZDC}$. In fact, it is possible to prove **9.1** using the Countable Principle of Choice $\mathbf{AC}_N$ instead of $\mathbf{DC}$, Problem ***x9.1**, but the proof is somewhat more technical. This is a general fact about the relation between $\mathbf{DC}$ and $\mathbf{AC}_N$: many results whose natural proofs call for $\mathbf{DC}$ follow from the weaker principle, with some additional effort.

Theorem **9.1** also settles the relation between infinite and Dedekind-infinite sets.

9.2. Corollary. *A set A is finite if and only if it is Dedekind-finite by **4.27**, i.e. if there exists no injection $\pi : A \rightarrow B \subsetneq A$ from A into one of its proper subsets.*

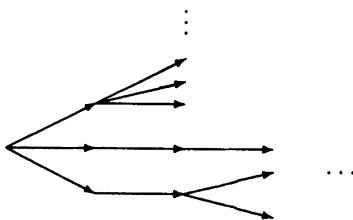


Figure 9.1. Non-empty tree.

Proof. Finite sets are Dedekind-finite by the Pigeonhole Principle, 5.25. If A is infinite, let $f : N \rightarrow A$ enumerate without repetitions some infinite, countable subset of it. The injection

$$\pi(x) = \begin{cases} f(n+1) & \text{if for some } n, x = f(n), \\ x & \text{if } x \notin f[N] \end{cases}$$

witnesses that A is Dedekind-infinite, since $\pi[A] = A \setminus \{f(0)\}$. $\dashv$

Next we consider an elementary but very useful result about trees, whose proof offers an additional illustration of the use of **DC** and its relation to **AC_N**.

9.3. Definition. A **tree**¹ on a set E is any set $T \subseteq E^*$ of strings from E which is closed under the relation of initial segment,

$$u \sqsubseteq v \in T \implies u \in T.$$

By (5.16), for strings, $u \sqsubseteq v \iff u \subseteq v$.

A lot of terms are used in the study of trees, most of them deriving from our picturing trees as, well, trees. The members of T are its **nodes** or **finite branches**, and every non-empty tree has $\emptyset$ as its least node, the **root**. If $u * \langle x \rangle \in T$, then u is a **parent** of $u * \langle x \rangle$ and $u * \langle x \rangle$ a **child** of u in T . Each node other than the root has exactly one parent, but may have many children; if it has none, it is a **terminal node**. With each node u we associate the **subtree**

$$T_u =_{\text{df}} \{w \in T \mid w \sqsubseteq u \vee u \sqsubseteq w\} \quad (9.1)$$

of nodes comparable with u . Easily,

$$T_u = \bigcup \{T_v \mid v \text{ is a child of } u\}. \quad (9.2)$$

¹Trees occur in many branches of mathematics, differently defined depending on the special needs of the field. The present definition is the most general we will need in these Notes.

9.4. Exercise. *Show (9.2).*

The **infinite branches** of a tree are its infinite sequences, and we collect them in the **body** of T ,

$$[T] =_{\text{df}} \{f : N \rightarrow E \mid (\forall n) \bar{f}(n) \in T\}. \quad (9.3)$$

Every infinite branch of a tree involves an infinite number of distinct nodes, so finite trees have empty bodies. It is also easy to construct infinite trees with empty bodies:

9.5. Exercise. *Show that the tree*

$$T = \{u \in N^* \mid (\forall i < lh(u), i > 0)[u(i) < u(i-1)]\}$$

on the natural numbers is infinite but has no infinite branch.

9.6. Definition. *A tree T is **finitely branching** if every node of T has at most finitely many children.* Notice that the tree in 9.5 is not finitely branching (at the root), and it could not be, by the following, basic result.

9.7. König's Lemma. *Every infinite, finitely branching tree has at least one infinite branch.*

Proof. Suppose $T \subseteq E^*$ is infinite, finitely branching, and let

$$S =_{\text{df}} \{u \in T \mid T_u \text{ is infinite}\}$$

be the subtree of those nodes in T which are comparable with infinitely many nodes. Since $T_\emptyset = T$ is infinite by hypothesis, the root $\emptyset \in S$, and (9.2) implies that

$$(\forall u \in S)(\exists v \in S)[v \text{ is a child of } u],$$

because each u has at most finitely many children and the infinite set S_u cannot be a finite union of finite sets. It follows by **DC** that there exists some $g : N \rightarrow S$ such that $g(0) = \emptyset$, and for every n , the value $g(n+1)$ is a child of $g(n)$, i.e. $g(n) \sqsubset g(n+1)$. Thus, the union

$$f = \bigcup \{g(n) \mid n \in N\}$$

is a (total) function $f : N \rightarrow E$, and for each n , $\bar{f}(n) = g(n)$, i.e. f is an infinite branch of T . $\dashv$

König's Lemma is very useful, especially in the following, more “constructive” version.

9.8. Definition. A set of nodes $B \subseteq T$ is a **bar** in a tree T , if every infinite branch of T passes through at least one node of B ,

$$(\forall f \in [T])(\exists n)\bar{f}(n) \in B.$$

9.9. Fan Theorem. If T is a finitely branching tree and B is a bar for T , then there exists a finite subset

$$B_0 = \{u_1, \dots, u_n\} \subseteq B$$

which is also a bar of T .

Proof. Let B_0 comprise the minimal members of the bar B ,

$$B_0 =_{\text{df}} \{u \in B \mid (\forall v \sqsubset u) v \notin B\},$$

and notice that B_0 is also a bar, because if $f \in [T]$ and n is least such that $\bar{f}(n) \in B$, then $\bar{f}(n) \in B_0$. Let S be the tree of all initial segments of the nodes in B_0 ,

$$S =_{\text{df}} \{v \in T \mid (\exists u \in B_0) v \sqsubseteq u\}.$$

Now S is a finitely branching tree (a subtree of T), and its terminal nodes are precisely the nodes in B_0 , because no member of B_0 is a proper initial segment of another. Thus, S cannot have an infinite branch, since B_0 is a bar for T . By König's lemma then, S is finite, and so its subset B_0 is also finite. $\dashv$

The surprisingly simple proof of König's Lemma is typical of arguments from **DC**, partly because its basic structure calls for **DC**, but also because of the following two reasons:

(1) König's Lemma can be proved for every tree T on a well orderable set E with no use of choice principles, Problem **x9.4**. In many applications, $E = N$ or E is finite, and then we need no choice whatsoever.

(2) Like **9.1**, König's Lemma can be proved by appealing to **AC_N** rather than **DC**, Problem ***x9.3**.

Many of the applications of the full Axiom of Choice have the following form: first we state and prove in **ZDC** (or even with no choice at all) some interesting proposition about well orderable sets, and then we infer the result we want for all sets by appealing to the Wellordering Theorem. Typical is the following generalization of the Hypothesis of Cardinal Comparability where (for the first and last time) we will state separately the corollary about all sets.

9.10. Theorem. Wellfoundedness of $\leq_c$. (1) For every non-empty class $\mathcal{E}$ of well orderable sets, there exists some $A_0 \in \mathcal{E}$ such that for every $A \in \mathcal{E}$, $A_0 \leq_c A$.

(2) **(AC)** Every non-empty class $\mathcal{E}$ of sets has a $\leq_c$ -least member.

Proof. By 7.33, let $U_0 = (A_0, \leq_0)$ be a $\leq_o$ -least well ordered set with field in $\mathcal{E}$. If $A \in \mathcal{E}$, then there exists some wellordering $\leq$ of A and by the choice of U_0 , $(A_0, \leq_0) \leq_o (A, \leq)$, so that, in particular, $A_0 \leq_c A$ since every initial similarity is an injection. $\dashv$

9.11. Lemma. The Next Cardinal. For every well orderable cardinal number κ , the cardinal

$$\kappa^+ =_{\text{df}} |\chi(\kappa)| \quad (9.4)$$

is also well orderable and it is least among the well orderable cardinals bigger than κ , i.e.

$$\kappa <_c \kappa^+, \quad \kappa <_c \lambda \implies \kappa^+ \leq_c \lambda, \quad (9.5)$$

for every well orderable cardinal λ . Here $\chi(\kappa)$ is the Hartogs well ordered set of κ , defined in 7.34.

Proof. Since κ^+ is well orderable, it is comparable with κ , it cannot be $\leq_c \kappa$ by Hartogs' Theorem 7.34, so $\kappa <_c \kappa^+$. The minimality of $\chi(\kappa)$ implies the rest. $\dashv$

We set

$$\aleph_1 =_{\text{df}} \aleph_0^+, \quad \aleph_2 =_{\text{df}} \aleph_1^+, \dots \quad (9.6)$$

9.12. Exercise. (AC) Since (with **AC**) every two cardinal numbers are comparable, the Continuum Hypothesis **CH** and the Generalized Continuum Hypothesis **GCH** can also be expressed by the simple identities

$$\mathbf{CH} \iff 2^{\aleph_0} =_c \aleph_1, \quad \mathbf{GCH} \iff (\forall \kappa)[2^\kappa =_c \kappa^+]. \quad (9.7)$$

This, unfortunately, does not help their resolution.

The next Lemma is often useful in arguments about well orderable sets.

9.13. Definition. A best wellordering of a set A is any wellordering $\leq$ of A in which every initial segment is smaller in cardinality than $|A|$,

$$(\forall x \in A)[|\text{seg}(x)| <_c |A|].$$

9.14. Lemma. (1) Every well orderable set admits a best wellordering.

(2) If $\leq_A, \leq_B$ are best wellorderings of A and B , then

$$A =_c B \implies (A, \leq_A) =_o (B, \leq_B).$$

In particular, any two best wellorderings of the same set are similar.

Proof. (1) Let $U = (A, \leq)$ be $\leq_o$ -least in the class of all well ordered sets with field A and suppose (towards a contradiction) that there exists some $x \in A$, $A \leq_c \mathbf{seg}_U(x)$. This yields an injection $\pi : A \rightarrow \mathbf{seg}_U(x)$ and the relation

$$u \leq' v \iff_{\text{df}} \pi(u) \leq \pi(v) \quad (u, v \in A)$$

is evidently a wellordering of A which is $\leq_o$ than $\mathbf{seg}_U(x)$ by 7.32, hence $<_o U$, contrary to the choice of U . (2) Suppose $U = (A, \leq_A)$, $V = (B, \leq_B)$ and (towards a contradiction) $U =_o \mathbf{seg}_V(x)$ for some $x \in B$. The similarity $\pi : A \rightarrow \mathbf{seg}_V(x)$ witnesses that

$$|A| =_c |\mathbf{seg}_V(x)| <_c |B|,$$

which is contrary to the hypothesis $|A| =_c |B|$. $\neg$

Every best wellordering of a countable, infinite set is similar with the natural wellordering of N , and we can use best wellorderings to show that many properties of countable sets hold for all well orderable sets. Typical is the next result, which generalizes the identity $\aleph_0^2 =_c \aleph_0$ and shows that the transfinite arithmetic of binary addition and multiplication is trivial.

9.15. Lemma. *For every infinite, well orderable set C , $C \times C =_c C$.*

Proof. Assume the contrary towards a contradiction, let C be a $\leq_c$ -least counterexample by 9.10 and let $\leq$ be a best wellordering of C . By the choice of C , for every infinite point $x \in C$,

$$|\mathbf{seg}(x)| + |\mathbf{seg}(x)| =_c 2 \cdot |\mathbf{seg}(x)| \leq_c |\mathbf{seg}(x)| \cdot |\mathbf{seg}(x)| <_c |C|. \quad (9.8)$$

The key step in the proof is the following definition of a new wellordering of the product $C \times C$, due to Gödel, which we have already met (somewhat concealed) in the proof of 5.31. We set

$$\begin{aligned} (x_1, y_1) \leq_g (x_2, y_2) \iff_{\text{df}} & [\max(x_1, y_1) < \max(x_2, y_2)] \\ & \vee [\max(x_1, y_1) = \max(x_2, y_2) \ \& \ x_1 < x_2] \\ & \vee [\max(x_1, y_1) = \max(x_2, y_2) \ \& \ x_1 = x_2 \\ & \quad \& \ y_1 \leq y_2]. \end{aligned} \quad (9.9)$$

The maximum here is obviously computed relative to the ordering $\leq$.

Lemma. *The relation $\leq_g$ is a wellordering of $C \times C$.*

Proof. For each non-empty $X \subseteq C \times C$, let w^* be $\leq$ -least such that for some $(x, y) \in X$, $\max(x, y) = w^*$; next let x^* be $\leq$ -least such that for some y , $(x^*, y) \in X$ and $\max(x^*, y) = w^*$; and finally let y^* be $\leq$ -least such that $(x^*, y^*) \in X$, $\max(x^*, y^*) = w^*$.

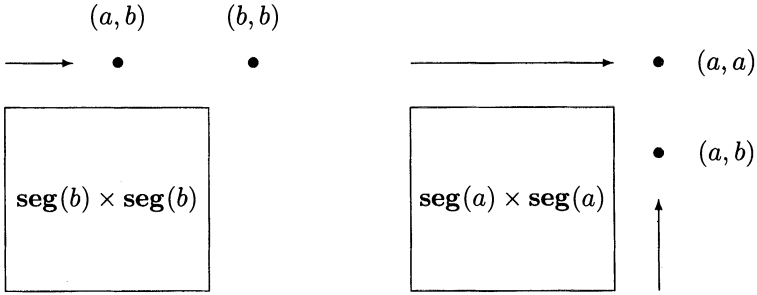


Figure 9.2. Initial segments of the Gödel wellordering.

The well ordered sets $(C, \leq)$ and $(C \times C, \leq_g)$ are $\leq_o$ -comparable and by the choice of $(C, \leq)$, $(C \times C, \leq_g) \leq_o (C, \leq)$ is not possible, so we must have $C <_o C \times C$; thus there exists some pair (a, b) of members of C such that

$$(C, \leq) =_o \text{seg}_{C \times C}((a, b)) = \text{seg}_g((a, b)),$$

and we will reach the desired contradiction if we can show that the initial segment $\text{seg}_g((a, b)) <_c C$. We consider the possibilities arising from the relative positions of a and b in $\leq$, and we use the fact that the point $\max(a, b)$ must be infinite.

CASE 1, $a = b$. From the definition of the Gödel wellordering,

$$(u, v) <_g (a, a) \iff [u < a \ \& \ v < a] \vee [u < a \ \& \ v = a] \vee [u = a \ \& \ v < a],$$

so that

$$\text{seg}_g((a, a)) = (\text{seg}(a) \times \text{seg}(a)) \cup (\text{seg}(a) \times \{a\}) \cup (\{a\} \times \text{seg}(a)),$$

and by repeated applications of (9.8),

$$|\text{seg}_g((a, a))| \leq_c |\text{seg}(a)|^2 + |\text{seg}(a)| \cdot 2 \leq_c |\text{seg}(a)| \cdot 3 <_c |C|.$$

CASE 2, $a < b$. Again from the definitions and because now $\max(a, b) = b$,

$$(u, v) <_g (a, b) \iff [u < b \ \& \ v < b] \vee [u < a \ \& \ v = b],$$

$\text{seg}((a, b)) = (\text{seg}(b) \times \text{seg}(b)) \cup (\text{seg}(a) \times \{b\})$ and a similar computation shows again that $|\text{seg}_g((a, b))| <_c |C|$.

CASE 3, $a > b$. This time

$$(u, v) <_g (a, b) \iff [u < a \ \& \ v < a] \vee [u < a \ \& \ v = a] \vee [u = a \ \& \ v < b],$$

from which we reach a contradiction as in the preceding cases. $\neg$

9.16. Corollary. The Absorption Laws. (AC) *If at least one of the cardinal numbers κ, λ is infinite and neither is 0, then*

$$\kappa + \lambda =_c \kappa \cdot \lambda =_c \max(\kappa, \lambda).$$

Proof. Assuming that $0 <_c \kappa \leq_c \lambda$ and using the result $\lambda \cdot \lambda =_c \lambda$ from the Lemma, we compute

$$\lambda \leq_c \kappa + \lambda \leq_c \kappa \cdot \lambda \leq_c \lambda \cdot \lambda =_c \lambda. \quad \dashv$$

9.17. Corollary. (AC) *For every indexed family of sets $(i \mapsto \kappa_i)_{i \in I}$ and every infinite κ , if $|I| \leq_c \kappa$ and for each $i \in I$, $\kappa_i \leq_c \kappa$, then $\sum_{i \in I} \kappa_i \leq_c \kappa$.*

Proof. Using **AC** and the hypothesis, choose for each $i \in I$ some injection $\pi_i : \kappa_i \rightarrow \kappa$, so that the mapping $((i, x) \mapsto (i, \pi_i(x)))$ is an injection of $\{(i, x) \mid i \in I \text{ \& } x \in \kappa_i\}$ into $I \times \kappa$. The existence of such an injection implies that

$$\sum_{i \in I} \kappa_i \leq_c |I \times \kappa| =_c |I| \cdot |\kappa| =_c \kappa. \quad \dashv$$

To find interesting problems and results in cardinal arithmetic we must consider operations with infinitely many arguments, of which the simplest are the following.

9.18. Cardinal Minimum Lemma. *There is a definite operation $\inf_c(\mathcal{E})$, such that for each non-empty family $\mathcal{E}$ of well orderable sets, the value $\kappa = \inf_c(\mathcal{E})$ has the following properties.*

1. κ is a well orderable cardinal number.
2. For some $A \in \mathcal{E}$, $\kappa =_c A$.
3. For every $B \in \mathcal{E}$, $\kappa \leq_c B$.

In addition, these conditions determine the value $\inf_c(\mathcal{E})$ up to $=_c$, i.e. if κ is any object which satisfies (1) - (3), then $\kappa =_c \inf_c(\mathcal{E})$.

Proof. If the cardinal assignment $|X|$ is strong by **4.21**, then **9.10** implies that there exists exactly one cardinal number which satisfies the condition

$$Least(\mathcal{E}, \kappa) \iff (\exists A \in \mathcal{E})[(\forall B \in \mathcal{E})[A \leq_c B] \text{ \& } \kappa = |A|],$$

and we can set

$$\inf_c(\mathcal{E}) = \text{the unique } \kappa \text{ such that } Least(\mathcal{E}, \kappa).$$

We need to do more, since we have only assumed that $|X|$ is a weak cardinal assignment and there may well exist many values of κ which satisfy $\text{Least}(\mathcal{E}, \kappa)$.

By Lemma 2 in the proof of Hartogs' Theorem 7.34, if $A \subseteq \bigcup E$ and $\leq$ is a wellordering of A , then the well ordered set $U = (A, \leq)$ is similar with some proper initial segment of $W = \chi(\bigcup \mathcal{E})$, and hence every $A \in \mathcal{E}$ is equinumerous with some proper initial segment of W . Thus we can set

$$\begin{aligned} w &=_{\text{df}} \text{ the least } x \in W \text{ such that } (\exists A \in \mathcal{E})[A =_c \text{seg}_W(x)], \\ \text{inf}_c(\mathcal{E}) &=_{\text{df}} |\text{seg}_W(w)|. \end{aligned}$$

Verification of the required properties of $\text{inf}_c(\mathcal{E})$ is quite trivial. ⊢

9.19. Exercise. *Show the part of the theorem which follows the “in addition.”*

9.20. Cardinal Supremum Lemma. *There is a definite operation $\text{sup}_c(\mathcal{E})$, such that for every non-empty family $\mathcal{E}$ of well orderable sets the set $\text{sup}_c(\mathcal{E})$ has the following properties.*

1. κ is a well orderable cardinal.
2. For every $A \in \mathcal{E}$, $A \leq_c \kappa$.
3. If B is well orderable and for all $A \in \mathcal{E}$, $A \leq_c B$, then $\kappa \leq_c B$.

In addition, these conditions determine the value $\text{sup}_c(\mathcal{E})$ up to $=_c$, i.e. if κ is any object which satisfies (1) - (3), then $\kappa =_c \text{inf}_c(\mathcal{E})$.

Proof. Let $C = h(\bigcup \mathcal{E})$ be the Hartogs set for the union of $\mathcal{E}$, which by Hartogs' Theorem 7.34 is well orderable and greater in cardinality than every well orderable subset of $\bigcup \mathcal{E}$, including every $A \in \mathcal{E}$. We set

$$\text{sup}_c(\mathcal{E}) =_{\text{df}} \text{inf}_c(\{B \subseteq C \mid (\forall A \in \mathcal{E})[A \leq_c B]\})$$

and verify easily the conclusion of the Lemma. ⊢

Infinite sums and products were defined in 4.21. We cannot say much about these, because infinite sums are as trivial as the finite ones (Problem x9.15), and infinite products are as complex as the Generalized Continuum Hypothesis, because

$$2^\kappa =_c \prod_{i \in \kappa} 2.$$

There is, however, a very interesting inequality relating the two.

9.21. König's Theorem. (AC) For any two families of sets $(i \mapsto A_i)$ and $(i \mapsto B_i)$ on the same index set $I \neq \emptyset$,

$$(\forall i \in I)[A_i <_c B_i] \implies \bigcup_{i \in I} A_i <_c \prod_{i \in I} B_i. \quad (9.10)$$

In particular, for families of cardinals, $(i \mapsto \kappa_i)$ and $(i \mapsto \lambda_i)$,

$$(\forall i \in I)[\kappa_i < \lambda_i] \implies \sum_{i \in I} \kappa_i < \prod_{i \in I} \lambda_i. \quad (9.11)$$

Proof. By the hypothesis and **AC**, for each i there exists an injection $\pi_i : A_i \hookrightarrow B_i$, and since π_i cannot be a bijection, there also exists a function $c : I \rightarrow \bigcup_{i \in I} B_i$ such that for each i , $c(i) \in B_i \setminus \pi_i[A_i]$. We set

$$\begin{aligned} f(x, i) &= \begin{cases} \pi_i(x), & \text{if } x \in A_i, \\ c(i), & \text{if } x \notin A_i, \end{cases} \\ g(x) &= (i \mapsto f(x, i)). \end{aligned}$$

If $x \neq y$ and x, y belong to the same A_i for some i , then $g(x)(i) = \pi_i(x) \neq \pi_i(y) = g(y)(i)$ because π_i is an injection, and hence $g(x) \neq g(y)$. If no A_i contains both x and y , suppose $x \in A_i$, $y \notin A_i$; it follows that $g(x)(i) = \pi_i(x) \in \pi_i[A_i]$ and $g(y)(i) = c(i) \in B_i \setminus \pi_i[A_i]$ so that again $g(x) \neq g(y)$. We conclude that the mapping $g : \bigcup_{i \in I} A_i \hookrightarrow \prod_{i \in I} B_i$ is an injection, and hence

$$\bigcup_{i \in I} A_i \leq_c \prod_{i \in I} B_i.$$

Suppose, towards a contradiction that there existed a correspondence

$$h : \bigcup_{i \in I} A_i \twoheadrightarrow \prod_{i \in I} B_i,$$

so that these two sets are equinumerous. For every i , the function

$$h_i(x) =_{\text{df}} h(x)(i) \quad (x \in A_i)$$

is (easily) an injection of A_i into B_i and by the hypothesis it cannot be a bijection; hence by **AC** there exists a function ε which selects in each B_i some element not in the image, i.e.

$$\varepsilon(i) \in B_i \setminus h_i[A_i], \quad (i \in I).$$

By its definition, $\varepsilon \in \prod_{i \in I} B_i$, so there must exist some $x \in A_j$, for some j , such that $h(x) = \varepsilon$; this yields

$$\varepsilon(j) = h(x)(j) = h_j(x) \in h_j[A_j],$$

contrary to the characteristic property of ε .

The cardinal version (9.11) follows by applying (9.10) to $A_i = \{i\} \times \kappa_i$ and $B_i = \lambda_i$. ⊣

9.22. Exercise. (AC) König's Theorem applies to the case $I = \kappa$, $\kappa_i = 1$ and $\lambda_i = 2$ and yields

$$\kappa = \bigcup_{i \in \kappa} \{i\} <_c \prod_{i \in \kappa} 2 =_c 2^\kappa,$$

i.e. the theorem of Cantor.

Despite its simplicity, König's Theorem implies immediately the only non-trivial inequality about the cardinal number $\mathfrak{c}$ of the continuum beyond Cantor's $\aleph_0 <_c \mathfrak{c}$, which is most naturally expressed using *cofinalities*.

9.23. Definition. The **cofinality** of a well orderable, infinite cardinal number κ is the least cardinal λ such that some sum of λ -many smaller than κ cardinals is equinumerous with κ :

$$cf(\kappa) =_{\text{df}} \inf_c(\{I \subseteq \kappa \mid \text{for some indexed family } (i \mapsto \kappa_i)_{i \in I}, \\ (\forall i \in I)[\kappa_i <_c \kappa] \text{ \& } \kappa =_c \sum_{i \in I} \kappa_i\}).$$

Notice that the family of well orderable index sets whose $\inf_c$ we take is not empty, it contains κ since

$$\kappa =_c \sum_{i \in \kappa} 1. \quad (9.12)$$

The general properties of $\inf_c$ imply the following basic properties of the cofinality operation:

1. $cf(\kappa) \leq_c \kappa$.
2. $\kappa =_c \sum_{i \in cf(\kappa)} \kappa_i$, for some $(i \mapsto \kappa_i)$ such that $(\forall i \in cf(\kappa))[\kappa_i <_c \kappa]$.
3. If λ is well orderable, $(\forall i \in \lambda)[\lambda_i <_c \kappa]$ and $\kappa =_c \sum_{i \in \lambda} \lambda_i$, then $cf(\kappa) \leq_c \lambda$.

Moreover, these conditions characterize $cf(\kappa)$ up to $=_c$.

A well orderable cardinal κ is **regular** if $cf(\kappa) =_c \kappa$, otherwise it is **singular**. It is convenient to define the operation $cf(\kappa)$ and the regularity condition for well orderable κ without assuming the full Axiom of Choice, but most results about these notions require **AC**.

9.24. Exercise. $\aleph_0$ is regular, because every finite sum of finite cardinals is finite.

9.25. Corollary. (AC) For each infinite cardinal number κ ,

$$cf(2^\kappa) >_c \kappa,$$

and in particular, $cf(\mathfrak{c}) >_c \aleph_0$, i.e. the continuum $\mathfrak{c}$ cannot be expressed as a countable sum of smaller cardinals.

Proof. By König's Theorem, if $\kappa_i <_c 2^\kappa$ for every $i \in \lambda$ with $\lambda \leq_c \kappa$, then

$$\begin{aligned} \sum_{i \in \lambda} \kappa_i &<_c \prod_{i \in \lambda} \kappa_i \leq_c \prod_{i \in \lambda} 2^\kappa \\ &=_{\text{c}} (2^\kappa)^\lambda =_{\text{c}} 2^{\kappa \cdot \lambda} =_{\text{c}} 2^\kappa, \end{aligned}$$

which contradicts $cf(2^\kappa) \leq_c \kappa$. ⊥

9.26. Gödel's model L of the constructible sets satisfies the Generalized Continuum Hypothesis, so for each κ , $2^\kappa =_c \kappa^+$ is regular by Problem **x9.17**. Using Cohen's forcing method, it is possible to construct models of **ZAC** in which $\mathfrak{c}$ is singular, with cofinality $cf(\mathfrak{c})$ some regular cardinal between $\aleph_0$ and $\mathfrak{c}$, for example $\aleph_1$.

We have left the basic properties of cofinalities for the problems, as they are very simple. We should remark, however, that it is not possible to study the topic seriously now, because without the Axiom of Replacement it is not possible to prove that singular cardinals *exist*!

Problems

***x9.1.** Show **9.1** using only the constructive axioms **(I)** - **(VI)** and the Countable Principle of Choice **AC_N**.

x9.2. Consider a system of airline routes which connects the (possibly infinitely many) cities of some world and assume the following. (1) From each city, there are only finitely many cities to which one can fly non-stop. (2) It is possible to travel by air from every city to every other city. (3) It is not possible to keep flying forever without visiting the same Airport twice. Show that this world has only finitely many cities.

***x9.3.** Show König's Lemma **9.7** using only the constructive axioms **(I)** - **(VI)** and the Countable Principle of Choice **AC_N**.

x9.4. Show König's Lemma **9.7** for the case where T is a tree on a well orderable set E , with no appeal to choice principles.

x9.5. Suppose T is a finitely splitting tree and B is a bar for T . Show that there exists some integer k , such that for all infinite branches $f \in [T]$, some $\bar{f}(i) \in B$, with $i \leq k$.

x9.6. Suppose C is a well orderable set, $f : C \times C \rightarrow C$, $A \subseteq C$, and let

$$A_f =_{\text{df}} \bigcap \{X \subseteq C \mid A \subseteq X \text{ \& } f[X \times X] \subseteq X\}$$

be the **closure** of A under f . Show that if A is infinite, then $A_f =_c A$.
HINT: Set by recursion $A_0 = A$, $A_{n+1} = A_n \cup f[A_n \times A_n]$ and show that $A_f = \bigcup_n A_n$.

x9.7. Show that if C is well orderable, then so is the set C^* of all words from C .

x9.8. If you used **AC_N** or **DC** in Problems **x9.6** and **x9.7**, do them again, using no choice principles at all.

x9.9. Every Hartogs set $h(A)$ is best wellordered by $\leq_{\chi(A)}$.

x9.10. If $(n \mapsto \kappa_n)_{n \in N}$ and $(n \mapsto \lambda_n)_{n \in N}$ are sequences of cardinal numbers, and for every n , $\kappa_n \leq_c \lambda_n$, then

$$\sum_{n \in N} \kappa_n \leq_c \sum_{n \in N} \lambda_n, \quad \prod_{n \in N} \kappa_n \leq_c \prod_{n \in N} \lambda_n.$$

x9.11. (AC) If $(i \mapsto \kappa_i)_{i \in I}$ and $(i \mapsto \lambda_i)_{i \in I}$ are families of cardinal numbers on the same index set I and for all $i \in I$, $\kappa_i \leq_c \lambda_i$, then

$$\sum_{i \in I} \kappa_i \leq_c \sum_{i \in I} \lambda_i, \quad \prod_{i \in I} \kappa_i \leq_c \prod_{i \in I} \lambda_i.$$

x9.12. (AC) For every indexed family of sets $(i \mapsto A_i)_{i \in I}$,

$$|\prod_{i \in I} A_i| =_c \prod_{i \in I} |A_i|$$

and the same for sums, with “disjoint union” on the left.

x9.13. (AC) Explain the notation and prove the identity

$$\prod_{i \in I} \prod_{j \in J(i)} \kappa_{ij} =_c \prod_{\{(i,j) \mid i \in I \text{ \& } j \in J(i)\}} \kappa_{ij}.$$

x9.14. (AC) Prove the characterization of $\sup_c(\mathcal{E})$ claimed in **9.20**.

x9.15. (AC) For every family of infinite cardinal numbers $(i \mapsto \kappa_i)$ on a non-empty index set I ,

$$\sum_{i \in I} \kappa_i =_c \max(|I|, \sup_c(\{\kappa_i \mid i \in I\})).$$

x9.16. (AC) Show that for every infinite cardinal κ , $cf(cf(\kappa)) =_c cf(\kappa)$, and hence $cf(\kappa)$ is always regular.

x9.17. (AC) For each infinite cardinal κ , the next cardinal κ^+ is regular.

x9.18. (AC) Show that for each infinite cardinal κ , $\kappa <_c \kappa^{cf(\kappa)}$.

***x9.19. (AC)** Every partial ordering $\leq$ on a set P has a **linearization**, i.e. some linear ordering $\leq'$ of P exists such that $x \leq y \implies x \leq' y$.

The next problem gives the basic fact which relates inductive and directed complete posets. Notice that by “chain” in a poset P we mean any subset $C \subseteq P$ which is linearly ordered by the poset partial ordering $\leq_P$; C is a *well ordered chain* if in addition, the restriction of $\leq_P$ to C is a wellordering. When we say that “ S is well orderable” for some $S \subseteq P$, we mean that S admits some wellordering $\leq$, which may be (and typically is) totally unrelated with the given partial ordering $\leq_P$ of P .

***x9.20.** If every well ordered chain in a poset $(P, \leq_P)$ has a least upper bound, then for every well orderable, directed subset S of P there exists a well ordered chain C with the following two properties.

(1) S is bounded by C , i.e. for each $x \in S$ there exists some $y \in C$ such that $x \leq_P y$.

(2) For each $y \in C$, there exists a directed subset $C_y \subseteq S$ such that $|C_y| <_c |S|$ and $y = \sup C_y$.

Notice that C may satisfy these conditions without being a subset of S . **HINT:** (W. Allen) Towards a contradiction, let S be well orderable, directed and $\leq_c$ -least counterexample to the conclusion, verify first that S must be uncountable, and let $\leq$ be a best wellordering of S . Define the function $f : S \times S \rightarrow S$ so that $x, y \in S \implies x, y \leq_P f(x, y)$, and for every $x \in S$, set

$$C_x =_{\text{df}} \text{seg}(x)_f,$$

with the notation of Problem **x9.6**. Show that this is directed, that $\sup C_x$ exists for each $x \in S$, and that

$$C =_{\text{df}} \{\sup C_x \mid x \in S\}$$

is a well ordered chain in P which has properties (1) and (2) for S .

***x9.21. (AC)** The following three conditions are equivalent, for every poset P :

1. Every directed set in P has a least upper bound.
2. Every chain in P has a least upper bound.
3. Every well ordered chain in P has a least upper bound.

In particular: **(AC)** A poset is inductive if and only if it is directed complete, a *dcpo*.

***x9.22. (AC)** Show that a monotone mapping $\pi : P \rightarrow Q$ on one inductive poset to another satisfies the identity

$$\pi(\sup S) = \sup \pi[S] \quad (9.13)$$

for every non-empty chain $S \subseteq P$, if and only if it satisfies (9.13) for every non-empty directed $S \subseteq P$.

x9.23. Prove that the characterization of continuity for mappings $\pi : (A \multimap E) \rightarrow (B \multimap M)$ in ***x6.22** holds for all A, E, B, M .

x9.24. (AC) Finite Basis Lemma. Let $\mathcal{J}$ be a non-empty family of subsets of some set V , such that

$$X \in \mathcal{J} \iff (\forall Y \subseteq X)[Y \text{ finite} \implies Y \in \mathcal{J}].$$

Show that $\mathcal{J}$ has a maximal member (under $\subseteq$).

***x9.25.** Let $\mathcal{J}$ be a family with the finite basis property as in **x9.24** and assume in addition that V is well orderable; show (without **AC**) that $\mathcal{J}$ has a maximal member.

x9.26. (AC) If you know what *vector spaces* are and the basic facts about *linear independence*, prove that every vector space has a basis. Prove also without **AC**, that every well orderable vector space has a basis. **HINT:** Apply **x9.24** or ***x9.25** to the family of all linearly independent subsets of the given space.

***x9.27. (AC)** If you know something about fields and algebraic extensions, prove that every field has an algebraic closure. **HINT:** The usual argument for this runs as follows. We consider the family

$$\mathcal{A} =_{\text{df}} \{F \mid F \text{ is an algebraic extension of } K\} \quad (9.14)$$

partially ordered by

$$F_1 \subseteq F_2 \iff_{\text{df}} F_1 \text{ is a subfield of } F_2,$$

we notice that it is an inductive poset, so that it has a maximal element $\overline{K}$, and we verify that this $\overline{K}$ is algebraically closed. The argument is defective, because the family $\mathcal{A}$ in (9.14) is not a set. To correct it, in the interesting case where K is infinite, we need to notice that *every algebraic extension of K is isomorphic with some field $F =_c K$* , so we can replace $\mathcal{A}$ in (9.14) by

$$\mathcal{A}' =_{\text{df}} \{F \subseteq E \mid F \text{ is an algebraic extension of } K\}, \quad (9.15)$$

where E is some superset of K with cardinality greater than K .

***x9.28.** Prove that every well orderable field (and in particular, every countable field) has an algebraic closure. **HINT:** The idea is to avoid use of **AC** by using Transfinite Recursion to construct the closure explicitly. You still need the trick suggested in the previous problem. Do the countable case first, it clarifies which algebraic results are needed.

Chapter 10

BAIRE SPACE

Next to the natural numbers, perhaps the most fundamental object of study of set theory is **Baire space**,

$$\mathcal{N} =_{\text{df}} (N \rightarrow N), \quad (10.1)$$

the set of all number theoretic sequences. If we let

$$\mathcal{C} =_{\text{df}} (N \rightarrow \{0, 1\}) \quad (10.2)$$

be the **Cantor set**¹ of all infinite, binary sequences, then $\mathcal{C} \subseteq \mathcal{N} \subseteq (N \times N)$ and from now familiar computations,

$$\mathfrak{c} =_c 2^{\aleph_0} =_c |\mathcal{P}(N)| =_c |\mathcal{C}| \leq_c |\mathcal{N}| \leq_c |\mathcal{P}(N \times N)| =_c |\mathcal{P}(N)| = \mathfrak{c}.$$

Since $\mathcal{N} =_c \mathcal{R}$ will follow as in Chapter 2 from the proper definitions in Appendix A, the Continuum Hypothesis **3.2** is equivalent to the proposition

$$(\mathbf{CH}) \quad (\forall X \subseteq \mathcal{N})[X \leq_c N \vee X =_c \mathcal{N}].$$

In fact, there is such a tight connection between $\mathcal{N}$, $\mathcal{C}$ and $\mathcal{R}$ that practically every interesting property of one of these spaces translates immediately to a related, interesting property of the others. In the problems we will make this precise for $\mathcal{N}$ and $\mathcal{C}$ and in Appendix A for $\mathcal{R}$, where we will also draw the consequences of the results of this chapter for the real numbers.²

The material in this chapter is not necessary for the comprehension of the two chapters which follow.

¹It is traditional to use the same name for this subset of $\mathcal{N}$ and the set of real numbers defined in the proof of **2.14**. Figure 2.4 explains vividly the reason for this and nobody has ever been confused by it.

²One may think of $\mathcal{N}$ as a “discrete,” “digital,” or “combinatorial” version of the “continuous” or “analog” $\mathcal{R}$. A real number x is completely determined by a decimal expansion $x(0).x(1)x(2)\dots$, where $(n \mapsto x(n)) \in \mathcal{N}$, but two distinct decimal expansions may compute to the same real number. This is a big “but”, it is the key fact behind the so-called *topological connectedness* of the real line which is of interest in analysis, to be sure, but of little set theoretic consequence. We may view Baire space as a “digital version” of $\mathcal{R}$ because it does not make any such identifications, each point $x \in \mathcal{N}$ determines unambiguously its “digits” $x(0), x(1), \dots$

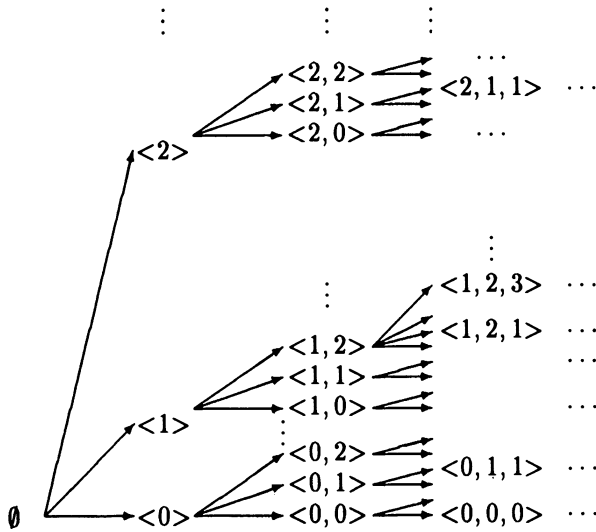


Figure 10.1. A small part of Baire space.

Our aim here is to establish some elementary facts about $\mathcal{N}$ which bear on the Continuum Problem. We will define the family of ANALYTIC SUBSETS OF $\mathcal{N}$ and prove that every analytic set satisfies the Continuum Hypothesis, in the sense that it must be either countable or equinumerous with $\mathcal{N}$. This PERFECT SET THEOREM 10.18 is significant because essentially every set of interest in classical analysis is analytic, including all the BOREL SETS which play a fundamental role in measure theory and integration. On the other hand, we will show in 10.26 that the basic and natural method of proving the Continuum Hypothesis for analytic sets cannot be extended to solve the full Continuum Problem, which remains open. In addition to their applications in analysis, these two results are of substantial foundational interest because their proofs illustrate beautifully the role of choice principles in classical mathematics.

10.1. The structure of $\mathcal{N}$. Our intuitions about $\mathcal{N}$ come from picturing it as the body of the largest tree on N in the terminology of 9.3 and (9.3),

$$\mathcal{N} = [N^*].$$

We will refer to subsets of Baire space as **pointsets**, the term “point” temporarily reserved for members of $\mathcal{N}$, infinite branches in N^* . By the **complement** of a pointset, we will mean its complement in $\mathcal{N}$,

$$cA =_{\text{df}} \mathcal{N} \setminus A. \quad (10.3)$$

It is convenient to extend the initial segment notation on strings,

$$u \sqsubseteq x \iff_{\text{df}} u \subset x \quad (u \in N^*, x \in \mathcal{N}), \quad (10.4)$$

to indicate that a finite sequence u is an initial segment of the point x , an **approximation** of x which determines the first $lh(u)$ values of x . For each $u \in N^*$, the set

$$\mathcal{N}_u =_{\text{df}} \{x \in \mathcal{N} \mid u \sqsubseteq x\} = [N_u^*] \quad (10.5)$$

of points in $\mathcal{N}$ which extend u is the **neighborhood** determined by u in $\mathcal{N}$.

10.2. Exercise. *The family of neighborhoods is countable.*

10.3. Definition. *A pointset A is **open** if it is a union of neighborhoods, so that*

$$x \in A \iff (\exists u)[x \in \mathcal{N}_u \text{ \& } \mathcal{N}_u \subseteq A];$$

*closed if its complement is open; and **clopen** if it is both closed and open.*

10.4. Exercise. *Every open pointset is the union of a sequence of neighborhoods.*

10.5. Proposition. (1) $\emptyset, \mathcal{N}$ and every neighborhood $\mathcal{N}_u$ are clopen. Every singleton $\{x\}$ is closed.

(2) The union $\bigcup \mathcal{G}$ of a family $\mathcal{G}$ of open pointsets is open and, dually, the intersection $\bigcap \mathcal{F}$ of a family $\mathcal{F}$ of closed pointsets is closed. We set $\bigcap \emptyset = \mathcal{N}$, so the intersection operation is defined for every family of pointsets.

(3) The intersection $G_1 \cap G_2$ of two open pointsets is open, and dually, the union $F_1 \cup F_2$ of two closed pointsets is closed.

Proof. These are all quite easy and we only give the proof of (3), as an example. If G_1, G_2 are open and $x \in G_1 \cap G_2$, then there exist $u, v \sqsubseteq x$ such that $\mathcal{N}_u \supseteq G_1$ and $\mathcal{N}_v \subseteq G_2$. The finite sequences u, v are comparable since they are both initial segments of x , so suppose $u \sqsubseteq v$, the argument being the same in the opposite case: now $\mathcal{N}_u \supseteq \mathcal{N}_v$, so $\mathcal{N}_v \subseteq G_1 \cap G_2$, as required. The dual property for closed sets follows by taking complements. $\dashv$

Baire space is a *topological space* by the classical definition recounted in 4.26, but a very special one, because of the next, basic connection between the topology and the combinatorial structure of the tree N^* . In proving it—and in the sequel, routinely—we will use the following trivial equivalence relating a tree T and its body:

$$x \in [T] \iff (\forall u \sqsubseteq x)[u \in T]. \quad (10.6)$$

It follows immediately from the definition of $[T]$, (9.3).

10.6. Proposition. *A pointset F is closed if and only if it is the body of a tree T on N , $F = [T]$.*

Proof. If $x \notin [T]$, then for some $u \sqsubseteq x$, $u \notin T$, and then $\mathcal{N}_u \cap [T] = \emptyset$, so $\mathcal{N}_u \subseteq c[T]$; thus $c[T]$ is open and $[T]$ is closed. Conversely, if we associate with each pointset F the tree

$$T^F =_{\text{df}} \{u \in N^* \mid (\exists x \in F)[u \sqsubseteq x]\}, \quad (10.7)$$

then obviously

$$F \subseteq [T^F].$$

If F is closed, we also have $[T^F] \subseteq F$: because if $x \notin F$, then for some $u \sqsubseteq x$, $\mathcal{N}_u \cap F = \emptyset$ by the openness of the complement cF , hence $u \notin T^F$ and $x \notin [T^F]$ by (10.6). $\dashv$

This basic characterization allows us to classify closed pointsets by the combinatorial properties of the trees which define them. It is not wrong to think of the cluster of combinatorial notions to come as the **geometry** of $\mathcal{N}$, although it is not a “geometry” by any standard, classical definition of this term.

10.7. Definition. Set

$$\begin{aligned} u \mid v &\iff_{\text{df}} u, v \text{ are incompatible} \\ &\iff (\exists i < lh(u), lh(v))[u(i) \neq v(i)], \end{aligned} \quad (10.8)$$

and, by extension,

$$u \mid x \iff_{\text{df}} \neg[u \sqsubseteq x] \iff (\exists v \sqsubseteq x)[u \mid v].$$

A string u **splits** in a tree T if it has incompatible extensions in T and a tree T is **splitting** if every $u \in T$ splits in T ,

$$u \in T \implies (\exists u_1, u_2 \in T)[u \sqsubseteq u_1 \ \& \ u \sqsubseteq u_2 \ \& \ u_1 \mid u_2].$$

A pointset P is **perfect** if it is the body of a splitting tree. Perfect sets are automatically closed.

10.8. Proposition. *Every non-empty, perfect pointset P had cardinality $\mathfrak{c}$.*

Proof. Suppose $P = [T]$ with T non-empty, splitting, and choose functions

$$l : T \rightarrow T, \quad r : T \rightarrow T$$

which witness the splitting property for T , i.e. for each $u \in T$,

$$u \sqsubseteq l(u), \quad u \sqsubseteq r(u), \quad l(u) \mid r(u).$$

By the Least Fixed Point Theorem **7.36** (or **6.21**), there exists a partial function

$$\sigma : \{0, 1\}^* \rightarrow T$$

from the tree of all binary strings into T which satisfies the identities

$$\sigma(\emptyset) = \emptyset, \quad \sigma(u * \langle 0 \rangle) = l(\sigma(u)), \quad \sigma(u * \langle 1 \rangle) = r(\sigma(u)),$$

and by an easy induction on $lh(u)$, σ is, in fact, total and $lh(u) \leq lh(\sigma(u))$. Moreover, $\sigma(u) \sqsubseteq \sigma(u * \langle i \rangle)$, for $i = 0, 1$, and another easy induction on $lh(v)$ establishes $\sigma(u) \sqsubseteq \sigma(u * v)$; this means that σ is monotone,

$$u \sqsubseteq v \implies \sigma(u) \sqsubseteq \sigma(v),$$

so we can define a function $\pi : \mathcal{C} \rightarrow [T]$ by

$$\pi(x) =_{\text{df}} \sup \{ \sigma(u) \mid u \sqsubseteq x \}. \quad (10.9)$$

The key property of σ is that it also preserves incompatibility,

$$u \mid v \implies \sigma(u) \mid \sigma(v). \quad (10.10)$$

To see this, let i be least such that $u(i) \neq v(i)$, so for some w we have

$$w * \langle 0 \rangle \sqsubseteq u, \quad w * \langle 1 \rangle \sqsubseteq v$$

(or the other way around); now $\sigma(w * \langle 0 \rangle)$ and $\sigma(w * \langle 1 \rangle)$ are incompatible and the monotonicity property implies that $\sigma(u), \sigma(v)$ extend them, so they are incompatible too. Finally, (10.10) implies that π is an injection, and this establishes that $\mathcal{C} \leq_c [T]$, which is all we need. $\dashv$

This simple abstraction of Cantor's proof of the uncountability of the reals (**2.14**) suggests an attack on the Continuum Problem: to prove that an uncountable pointset has cardinality $\mathfrak{c}$, it is enough to show that it contains a non-empty, perfect subset. This is trivially true of open sets (because each $\mathcal{N}_u$ is perfect) and it is also true of closed sets, less trivially.

10.9. Cantor-Bendixson Theorem. *Every closed subset F of $\mathcal{N}$ can be decomposed uniquely into two disjoint subsets*

$$F = P \cup S, \quad P \cap S = \emptyset, \quad (10.11)$$

where P , the **kernel** of F , is perfect and S , the **scattered part** of F , is countable. It follows that every uncountable, closed pointset has a non-empty, perfect kernel and hence has cardinality $\mathfrak{c}$.

Proof. Let $T = T^F$ as in (10.7), so that T has no terminal nodes and $F = [T]$, and set

$$\begin{aligned} S &=_{\text{df}} \bigcup \{ [T_u] \mid u \in T \text{ \& } |[T_u]| \leq_c \aleph_0 \}, \\ P &=_{\text{df}} F \setminus S. \end{aligned}$$

By its definition, S is the union of countably many, countable sets, so it is countable (note the use of $\mathbf{AC}_N$ here), and it remains to show that P is perfect.

The set of strings

$$kT = \{u \in T \mid |[T_u]| >_c \aleph_0\}$$

is easily a tree, and

$$x \in S \iff x \in F \ \& \ (\exists u \sqsubseteq x)[u \notin kT]$$

is another way to read the definition of S . Since $P = F \setminus S$,

$$\begin{aligned} x \in P &\iff x \in F \ \& \ [x \notin F \vee (\forall u \sqsubseteq x)[u \in kT]] \\ &\iff (\forall u \sqsubseteq x)[u \in T] \ \& \ (\forall u \sqsubseteq x)[u \in kT] \\ &\iff (\forall u \sqsubseteq x)[u \in kT] \\ &\iff x \in [kT], \end{aligned}$$

and it is enough to prove that kT is splitting. Suppose, towards a contradiction that some $u \in kT$ does not split. This means that all extensions of u in kT are compatible and they define a single point

$$x = \sup \{v \in kT \mid u \sqsubseteq v\}.$$

Since every extension of u in kT approximates x ,

$$[T_u] = \{x\} \cup \bigcup \{[T_v] \mid u \sqsubseteq v \in T \ \& \ |[T_v]| \leq_c \aleph_0\};$$

this, however, implies that $[T_u]$ is a countable union of countable sets, which is absurd.

We leave the uniqueness of the decomposition (10.11) for the problems, **x10.1**. ⊣

10.10. Definition. A family Γ of pointsets has **property P** if every uncountable set in Γ contains a non-empty, perfect subset. In this classical terminology,³ the family $\mathcal{F}$ of closed pointsets has property **P**, or (more simply) every closed pointset has property **P**.

³The classical terminology in question is quite absurd, but so well established that it would be folly to change it or bypass it. In any topological space, closed sets are $\mathcal{F}$ -sets, from the French *fermèt*, open sets are $\mathcal{G}$ -sets, from the German *Gebiete* (it means *region*), countable unions of Γ -sets are Γ_σ -sets and countable intersections of Γ -sets are Γ_δ -sets, from the German words *Summe* and *Durchschnitt* for *union* and *intersection*, respectively. We will only use this terminology in passing references to $\mathcal{F}_\sigma$ and $\mathcal{G}_\delta$ pointsets.

10.11. Exercise. If a family of pointsets Γ has property **P**, then the family Γ_σ of all countable unions of sets in Γ also has property **P**.

Thus every $\mathcal{F}_\sigma$ pointset, of the form

$$A = \bigcup_{n \in \mathbb{N}} F_n \quad (10.12)$$

with each F_n closed has property **P**. The same is true of every $\mathcal{G}_\delta$ set, of the form

$$A = \bigcap_{n \in \mathbb{N}} G_n \quad (10.13)$$

with each G_n open, but the proof is not that simple. It is best to jump directly to a much bigger family of pointsets, defined in such a way that proving property **P** for them is simple. The program is to define first the class of analytic pointsets, show that it has property **P**, and then establish some strong closure properties for it which will give us a wealth of pointsets with property **P**.

10.12. Definition. Recall from 6.24 that a function $f : X \rightarrow Y$ on one topological space to another is **continuous** if the inverse image $f^{-1}[G]$ of every open set in Y is open in X . A pointset⁴ $A \subseteq \mathcal{N}$ is **analytic** or **Suslin**, if either $A = \emptyset$ or A is the image of Baire space under a continuous function, in symbols,

$$\mathcal{A} =_{\text{df}} \{A \subseteq \mathcal{N} \mid A = \emptyset \vee (\exists \text{ continuous } f : \mathcal{N} \rightarrow \mathcal{N})[A = f[\mathcal{N}]]\}.$$

Continuity in $\mathcal{N}$ has a simple, combinatorial interpretation which is the key to its applications.

10.13. Theorem. A function $f : \mathcal{N} \rightarrow_{\tau} \mathcal{N}$ is continuous if and only if there exists a monotone function $\tau : N^* \rightarrow N^*$ on strings, such that

$$\begin{aligned} f(x) &= \sup \{\tau(u) \mid u \sqsubseteq x\} \quad (x \in \mathcal{N}) \\ &= \lim_n \tau(\bar{x}(n)). \end{aligned} \quad (10.14)$$

When $\tau : N^* \rightarrow N^*$ is monotone and (10.14) holds, we say that τ **computes** the function f .

Proof. If f satisfies (10.14), then

$$f(x) \in \mathcal{N}_v \iff (\exists u \sqsubseteq x)[v \sqsubseteq \tau(u)],$$

⁴We stick to Baire space here because there are many competing, inequivalent definitions of “analytic sets” which are not equivalent in general topological spaces, although they all agree on $\mathcal{N}$.

so each inverse image of a neighborhood

$$f^{-1}[\mathcal{N}_v] = \bigcup \{\mathcal{N}_u \mid v \sqsubseteq \tau(u)\}$$

is a union of neighborhoods and f is continuous. For the converse, suppose f is continuous and let

$$S(u) =_{\text{df}} \{v \in N^* \mid f[\mathcal{N}_u] \subseteq \mathcal{N}_v\} \quad (u \in N^*).$$

Each $S(u) \neq \emptyset$, since the root $\emptyset \in S(u)$, $v \sqsubseteq v' \in S(u) \implies v \in S(u)$, and

$$v, v' \in S(u) \implies f[\mathcal{N}_u] \subseteq \mathcal{N}_v \cap \mathcal{N}_{v'} \implies [v \sqsubseteq v' \vee v' \sqsubseteq v],$$

since $v \vee v' \implies \mathcal{N}_v \cap \mathcal{N}_{v'} = \emptyset$. Thus, there are two possibilities.

CASE 1. There is some $v \in S(u)$ such that $lh(v) = lh(u)$. In this case we set

$$\tau(u) =_{\text{df}} v = \text{the unique string in } S(u) \text{ such that } lh(v) = lh(u).$$

CASE 2. There is no $v \in S(u)$ such that $lh(v) = lh(u)$. Now we set

$$\tau(u) =_{\text{df}} \sup \{v \mid v \in S(u)\}.$$

The monotonicity of τ follows easily from the implications

$$u_1 \sqsubseteq u_2 \implies f[\mathcal{N}_{u_1}] \supseteq f[\mathcal{N}_{u_2}] \implies S(u_1) \subseteq S(u_2),$$

considering the possibilities in the definitions of $\tau(u_1)$ and $\tau(u_2)$. To prove (10.14), notice first that because $\tau(u) \in S(u)$,

$$u \sqsubseteq x \in \mathcal{N} \implies f(x) \in \mathcal{N}_{\tau(u)} \implies \tau(u) \sqsubseteq f(x).$$

Moreover, by the continuity of f , if $v \sqsubseteq f(x)$, then for some $u \sqsubseteq x$, $f[\mathcal{N}_u] \subseteq \mathcal{N}_v$, hence $v \in S(u)$ and either immediately $v \sqsubseteq \tau(u)$, if $\tau(u)$ is defined by CASE 2, or there is some u' extending u , with $lh(u') = lh(v)$ such that $v = \tau(u')$ in the other case. $\dashv$

It is useful to think of (10.14) as a computational characterization of continuity: the string function $\tau(u)$ gives us better and better approximations $\tau(u) \sqsubseteq f(x)$ to the value of f , as we feed into it successively finer approximations $u \sqsubseteq x$ to the argument. We can turn this picture into a precise and elegant result, in terms of the notions introduced in Chapter 6.

10.14. Corollary. *A function $f : \mathcal{N} \rightarrow \mathcal{N}$ is continuous if and only if it is the restriction to $\mathcal{N}$ of some monotone, continuous mapping*

$$\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$$

on the inductive poset $(N \rightarrow N)$. By Definition 6.22, a monotone mapping $\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$ is continuous if it satisfies the equivalence

$$\pi(x)(i) = w \iff (\exists u \in N^*)[u \sqsubseteq x \ \& \ \pi(u)(i) = w].$$

Here we are using the fact that $\mathcal{N}$ as a subset of the inductive poset $(N \rightarrow N)$, consisting precisely of all the maximal elements of $(N \rightarrow N)$. The basic observation is the decomposition

$$(N \rightarrow N) = N^* \cup \mathcal{N}, \quad N^* \cap \mathcal{N} = \emptyset. \quad (10.15)$$

Proof. If $f : \mathcal{N} \rightarrow \mathcal{N}$ is continuous, let τ compute it by the Theorem and take (literally)

$$\pi = \tau \cup f,$$

i.e. $\pi(u) = \tau(u)$ for $u \in N^*$ and $\pi(x) = f(x)$ for $x \in \mathcal{N}$. The continuity of π is trivial. The converse is very easy. $\dashv$

10.15. Exercise. Prove the “easy converse,” i.e. that if $f : \mathcal{N} \rightarrow \mathcal{N}$ is the restriction to $\mathcal{N}$ of some continuous $\pi : (N \rightarrow N) \rightarrow (N \rightarrow N)$, then f is continuous.

The Corollary makes it possible to recognize continuity of specific functions on Baire space instantly, by inspection, simply noticing that every digit $f(x)(i)$ of each value $f(x)$ can be computed using only finitely many values of x . As in Chapter 6, a passing remark that some function or other is “evidently continuous” accompanied by no proof typically means an appeal to this result.

10.16. Definition. A pointset K is **compact** if $K = [T]$ is the body of a finitely branching tree T on N . In particular, every compact pointset is closed and $\mathcal{C}$ is compact.

Some cheating is involved in adopting this as the definition of compactness for pointsets, since there is a perfectly general definition of compactness for sets in arbitrary topological spaces, by which 10.16 is a theorem. Without comment, we did the same for perfection, which is also a general, topological notion. What we need here are the combinatorial properties of these pointsets specific to Baire space and we have relegated their topological characterizations to the problems, **x10.16** and ***x10.20**.

10.17. Proposition. (1) The image $f[K]$ of a compact pointset K by a continuous function $f : \mathcal{N} \rightarrow \mathcal{N}$ is compact.

(2) The image $f[K]$ of a compact and perfect pointset K by a continuous injection $f : \mathcal{N} \rightarrow \mathcal{N}$ is compact and perfect.

Proof. (1) Suppose $K = [T]$ where T is finitely branching, τ computes f as in (10.14), and let

$$S = T^{f[K]} = \{v \mid (\exists x \in K)[v \sqsubseteq f(x)]\}$$

be the tree of initial segments of the image $f[K]$. It is enough to prove that S is a finitely branching tree and $f[K] = [S]$.

To see first that S is finitely branching, suppose $v \in S$, let

$$B =_{\text{df}} \{u \in T \mid v \mid \tau(u) \vee v \sqsubseteq \tau(u)\},$$

and suppose that $x \in [T]$. If $v \mid f(x)$, then for some n , $v \mid \tau(\bar{x}(n))$, so $\bar{x}(n) \in B$; and if $v \sqsubseteq f(x)$, then for some n , $v \sqsubseteq \tau(\bar{x}(n))$, and again $\bar{n} \in B$. Thus, B is a bar for T , and by the Fan Theorem 9.9 it must have a finite subset

$$B_0 = \{u_0, \dots, u_n\} \subseteq B$$

which is also a bar. Thus, for every $x \in K$ such that $v \sqsubseteq f(x)$, there exists some u_i such that $v \sqsubseteq \tau(u_i) \sqsubseteq f(x)$, so that every child of v in S is an initial segment of some $\tau(u_i)$, and there are only finitely many of those.

Clearly, $f[K] \subseteq [S]$. To prove $[S] \subseteq f[K]$, suppose towards a contradiction that $y \in [S] \setminus f[K]$ and let

$$B =_{\text{df}} \{u \in T \mid \tau(u) \mid y\}.$$

Now B is a bar for T , because the only way that $\tau(\bar{x}(n))$ can be compatible with y for every n is if $f(x) = y$. By the Fan Theorem again, there is a finite subset

$$B_0 = \{u_0, \dots, u_n\} \subseteq B$$

which is also a bar for T . Let

$$k = \max\{lh(\tau(u_i)) \mid i \leq n\} + 1,$$

and choose some $x \in [T]$ such that $\bar{y}(k) \subseteq f(x)$, which exists because $y \in [S]$, so that y can be approximated arbitrarily well by points in the image of f . On the other hand, $u_i \sqsubseteq x$ for some i since B_0 is a bar; hence $\tau(u_i) \sqsubseteq f(x)$ because τ computes f ; so both $\tau(u_i)$ and $\bar{y}(k)$ are initial segments of $f(x)$, and hence compatible; and since $\tau(u_i)$ has smaller length than $\bar{y}(k)$, this means that $\tau(u_i) \sqsubseteq y$, which contradicts the definition of B .

(2) With the same notation as in (1) and the additional hypothesis, let $v \in S$, so that for some $u \in T$, $v \sqsubseteq \tau(u)$. Since T is splitting, there exist distinct points

$$x_1, x_2 \in K \cap \mathcal{N}_u,$$

and since τ computes f ,

$$\tau(u) \sqsubseteq f(x_1), \quad \tau(u) \sqsubseteq f(x_2). \quad (10.16)$$

But $f(x_1) \neq f(x_2)$, because f is an injection, so there exist incompatible $v_1 \sqsubseteq f(x_1)$, $v_2 \sqsubseteq f(x_2)$, which extend $\tau(u)$ by (10.16), so they split $\tau(u)$ and the smaller v in S . $\dashv$

10.18. Perfect Set Theorem (Suslin, 1916). *Every uncountable, analytic set has a non-empty, perfect subset.*

Proof. Assume that $A = f[N^*]$ is uncountable, suppose τ computes f , and let

$$T =_{\text{df}} \{u \in N^* \mid |f[\mathcal{N}_u]| >_c \aleph_0\}. \quad (10.17)$$

Clearly, T is a non-empty tree.

Lemma. The tree T is τ -splitting, i.e. for each $u \in T$, there exist $u_1, u_2 \in T$ such that

$$u \sqsubseteq u_1, \quad u \sqsubseteq u_2, \quad \tau(u_1) \mid \tau(u_2).$$

Proof. For any $u \in T$ and any fixed $x \in \mathcal{N}_u$,

$$f[\mathcal{N}_u] = \{f(x)\} \cup \bigcup \{f[\mathcal{N}_{u'}] \mid \tau(u') \mid f(x)\} \quad (10.18)$$

since $f(y) \neq f(x) \implies \tau(u') \sqsubseteq f(y)$ for some u' such that $\tau(u')$ is incompatible with $f(x)$. If the Lemma fails at u , then

$$u \sqsubseteq u' \in T \implies \tau(u') \sqsubseteq f(x);$$

thus each image $f[\mathcal{N}_{u'}]$ with $\tau(u') \mid f(x)$ in (10.18) involves some $u' \notin T$ and is countable, and there are only countably many choices for u' . Thus, $f[\mathcal{N}_u]$ is the union of a singleton and a countable family of countable sets, hence countable, contrary to hypothesis.

As in 10.8, we choose functions

$$l : T \rightarrow T, \quad r : T \rightarrow T$$

which witness the τ -splitting property for T , i.e. for each $u \in T$,

$$u \sqsubseteq l(u), \quad u \sqsubseteq r(u), \quad \tau(l(u)) \mid \tau(r(u)),$$

and we define by 7.36 or 6.21 a partial function

$$\sigma : \{0, 1\}^* \rightarrow T$$

from the tree of all binary strings into T which satisfies the identities

$$\sigma(\emptyset) = \emptyset, \quad \sigma(u * \langle 0 \rangle) = l(\sigma(u)), \quad \sigma(u * \langle 1 \rangle) = r(\sigma(u)),$$

and which, as a consequence, is total and monotone. The key property of this σ is that it takes incompatible binary strings to τ -incompatible strings,

$$u \mid v \implies \tau(\sigma(u)) \mid \tau(\sigma(v)), \quad (10.19)$$

and it is verified exactly as (10.10) was verified in the proof of **10.8**. In addition, σ computes a continuous $g : \mathcal{C} \rightarrow \mathcal{N}$,

$$g(x) = \sup \{ \sigma(u) \mid u \sqsubseteq x \},$$

and evidently

$$g[\mathcal{C}] \subseteq [T]. \quad (10.20)$$

Consider now the composition $h = fg$ of the given f and this g , which is computed by the composition of τ and σ :

$$h(x) = \sup \{ \tau(\sigma(u)) \mid u \sqsubseteq x \}.$$

This is continuous and injective by (10.19), so its image $h[\mathcal{C}] = fg[\mathcal{C}]$ is compact and perfect by **10.17**, and it is included in $f[T] \subseteq A$ by (10.20). $\dashv$

The result means nothing, of course, until we prove that there are lots and lots of analytic sets.

10.19. Lemma. *Every closed pointset is analytic.*

Proof. Let $T = T^F$ as in (10.7) for the given closed set $F \neq \emptyset$, so in addition to $F = [T]$ we also know that every string in T has an extension, there are no terminal nodes. Thus, we can fix a function $l : T \rightarrow T$ such that

$$u \in T \implies u \sqsubseteq l(u) \text{ \& } lh(l(u)) = lh(u) + 1.$$

Let also

$$rtail(u) = \bar{u}(lh(u) - 1) \quad (lh(u) > 0) \quad (10.21)$$

be the partial function which strips each non-empty string of its last element. By **7.36** or **6.21**, there exists a partial function $\tau : N^* \rightarrow T$ such that

$$\tau(u) = \begin{cases} u, & \text{if } u \in T, \\ l(\tau(rtail(u))), & \text{if } u \notin T, \end{cases}$$

which is (easily) a *projection* of N^* onto T , i.e. it is total, length preserving and the identity on T , and which (as a consequence) computes a function $f : \mathcal{N} \rightarrow [T]$. $\dashv$

10.20. Lemma. *Every continuous image of an analytic pointset is analytic.*

Proof. If $A = f[B]$ and $B = g[\mathcal{N}]$, then $A = fg[\mathcal{N}]$, and the composition fg is continuous. $\dashv$

10.21. Lemma. *If $f, g : \mathcal{N} \rightarrow \mathcal{N}$ are continuous functions, then the set*

$$E = \{x \mid f(x) = g(x)\}$$

of points on which they agree is closed.

Proof. Because distinct points can be approximated by incompatible initial segments,

$$\begin{aligned} x \notin E &\iff f(x) \neq g(x) \\ &\iff (\exists u, v)[f(x) \in \mathcal{N}_u \ \& \ g(x) \in \mathcal{N}_v \ \& \ u \perp v], \end{aligned}$$

which means that

$$cE = \bigcup \{f^{-1}[\mathcal{N}_u] \cap g^{-1}[\mathcal{N}_v] \mid u \perp v\},$$

so that cE is the union of open sets and hence open. $\dashv$

10.22. Theorem. *Countable unions and countable intersections of analytic pointsets are analytic.*

Proof. Suppose that $A_n = f_n[\mathcal{N}]$ with each f_n continuous and define first $f : \mathcal{N} \rightarrow \mathcal{N}$ by the formula

$$f(z) = f_{z(0)}(\text{tail}(z)),$$

where

$$\text{tail}(z) = (i \mapsto z(i+1)) = (z(1), z(2), \dots)$$

is the function which decapitates points. Evidently f is continuous: because each $f(z)(i)$ can be computed from finitely many values of z , first setting $n = z(0)$ and then using the finitely many values of $\text{tail}(z)$ needed to compute $f_n(\text{tail}(z))$. Moreover:

$$\begin{aligned} y \in \bigcup_n f_n[\mathcal{N}] &\iff (\exists n \in N, x \in \mathcal{N})[y = f_n(x)] \\ &\iff (\exists z \in \mathcal{N})[y = f_{z(0)}(\text{tail}(z))] \\ &\quad \text{taking } z(0) = n, \text{tail}(z) = x \\ &\iff (\exists z \in \mathcal{N})[y = f(z)], \end{aligned}$$

so $\bigcup_n A_n = f[\mathcal{N}]$ and the union of the A_n 's is analytic.

The key fact for this argument was that the mapping

$$z \mapsto (z(0), \text{tail}(z))$$

is a surjection of $\mathcal{N}$ onto $N \times \mathcal{N}$ —actually a bijection—with continuous components. To prove that the intersection $\bigcap_n A_n$ is analytic, we need a similar surjection

$$\pi : \mathcal{N} \rightarrow (N \rightarrow \mathcal{N})$$

of $\mathcal{N}$ onto the set of infinite sequences of points. Fix some bijection $\rho : N \times N \rightarrow N$ and set

$$\rho_n(z) = (i \mapsto z(\rho(n, i))). \quad (10.22)$$

Each $\rho_n : \mathcal{N} \rightarrow \mathcal{N}$ is clearly continuous and for each infinite sequence $\{x_n\}_{n \in N}$ of points we can define z such that

$$z(\rho(n, i)) = x_n(i) \quad (n, i \in N),$$

so that

$$\rho_n(z) = x_n \quad (n \in N);$$

in other words, the mapping

$$\pi(z) = (n \mapsto \rho_n(z))$$

is a surjection. Using **AC**_N now,

$$\begin{aligned} y \in \bigcap_n A_n &\iff (\forall n)(\exists x)[y = f_n(x)] \\ &\iff (\exists \{x_n\}_{n \in N})(\forall n)[y = f_n(x_n)] \\ &\iff (\exists z \in \mathcal{N})(\forall n)[y = f_n(\rho_n(z))] \\ &\iff (\exists z \in \mathcal{N})[(\forall n)[f_n(\rho_n(z)) = f_0(\rho_0(z))] \\ &\quad \& y = f_0(\rho_0(z))]. \end{aligned} \quad (10.23)$$

For each n , the set

$$B_n = \{z \in \mathcal{N} \mid f_n(\rho_n(z)) = f_0(\rho_0(z))\}$$

is closed by **10.21**, hence the intersection

$$B = \bigcap_n B_n$$

is also closed. From (10.23), however,

$$\bigcap_n A_n = f_0 \rho_0[B],$$

which means that the intersection of the A_n 's is analytic. $\dashv$

10.23. Definition. The family $\mathcal{B}(X)$ of the **Borel** subsets of a topological space X is the smallest family of subsets of X which includes the open sets and is a σ -field, i.e. it is closed under countable unions and complementation:

$$\begin{aligned} (\forall n)[A_n \in \mathcal{B}(X)] &\implies \bigcup_n A_n \in \mathcal{B}(X), \\ A \in \mathcal{B}(X) &\implies cA \in \mathcal{B}(X). \end{aligned}$$

We are mostly interested in Baire space of course,

$$\mathcal{B} =_{\text{df}} \mathcal{B}(\mathcal{N}) = \text{the family of Borel pointsets.}$$

10.24. Exercise. *Prove that the definition makes sense, i.e. the intersection*

$$\begin{aligned} \mathcal{B}(X) = \bigcap \{ \mathcal{E} \mid & \mathcal{G} \subseteq \mathcal{E} \\ & \& (\forall \{A_n\}_n \subseteq \mathcal{E}) [\bigcup_n A_n \in \mathcal{E}] \\ & \& (\forall A \in \mathcal{E}) [cA \in \mathcal{E}] \} \end{aligned}$$

is a σ -field which contains the open sets, and hence the least such.

10.25. Corollary. *Every Borel pointset is analytic (Suslin) and hence has property $\mathbf{P}$ (Alexandroff, Hausdorff).*

Proof. Let

$$C\mathcal{A} = \{A \subseteq \mathcal{N} \mid cA \in \mathcal{A}\} \quad (10.24)$$

be the family of **co-analytic pointsets**, those with analytic complements. The family $\mathcal{A} \cap C\mathcal{A}$ of pointsets which are both analytic and co-analytic is a σ -field, since it is closed under complementation by definition, and if each $A_n \in \mathcal{A} \cap C\mathcal{A}$, then $\bigcup_n A_n$ and

$$c(\bigcup_n A_n) = \bigcap_n cA_n$$

are both analytic by the theorem. Since every open set

$$G = \bigcup_n \{\mathcal{N}_u \mid \mathcal{N}_u \subseteq G\}$$

is a countable union of neighborhoods, hence analytic, and also co-analytic by **10.19**, $\mathcal{A} \cap C\mathcal{A}$ is a σ -field which contains all the open pointsets and hence includes every Borel set. $\dashv$

Suslin introduced the family of analytic pointsets in 1917 and proved a slew of theorems about it, including his famous characterization

$$\mathcal{A} \cap C\mathcal{A} = \mathcal{B}. \quad (10.25)$$

He also showed that not every analytic pointset is an analytic complement, so the inclusion $\mathcal{B} \subsetneq \mathcal{A}$ is proper. The Borel sets had been introduced more than a decade earlier by Borel and Lebesgue and they were the key to the successful development of the theory of *Lebesgue integration*, one of the chief achievements of 19th century analysis. For most purposes of integration theory, including its later, fundamental applications to *probability*, every pointset of interest is *almost equal* to a Borel set, in a precise sense which basically allows us to study the subject as if every pointset were Borel. Because of this, the special case of the Continuum Problem for Borel sets was thought very important and its simultaneous, independent solutions published by Alexandroff and Hausdorff in 1916 brought

instant recognition to those two, then very young and later very famous mathematicians.

The family of analytic sets falls far short from exhausting the powerset of $\mathcal{N}$, Problem **x10.8**. Still, one might hope that the method we used to solve the Continuum Problem for them might be extended to prove the full Continuum Hypothesis, but this too is far from the mark.

10.26. Theorem. (AC) *There exists a pointset $A \subset \mathcal{N}$ which is uncountable but contains no non-empty perfect set.*

Proof. The key fact is that there are exactly as many non-empty, perfect sets as there are points in $\mathcal{N}$:

Lemma 1. If $\mathcal{P} = \{P \subseteq \mathcal{N} \mid P \neq \emptyset, P \text{ perfect}\}$, then $|\mathcal{P}| =_{\mathfrak{c}} \mathfrak{c}$.

Proof. For each $y \in \mathcal{N}$, the pointset

$$A_y = \{x \mid (\forall n)[y(n) \leq x(n)]\}$$

is easily perfect, equally easily $y \neq z \implies A_y \neq A_z$, hence $\mathfrak{c} =_{\mathfrak{c}} |\mathcal{N}| \leq_{\mathfrak{c}} |\mathcal{P}|$. On the other hand, each perfect set $P = [T^P]$ is the body of a tree on N which determines it, so

$$|\mathcal{P}| \leq_{\mathfrak{c}} |\mathcal{P}(N^*)| =_{\mathfrak{c}} |\mathcal{P}(N)| =_{\mathfrak{c}} \mathfrak{c}.$$

Fix a set

$$I =_{\mathfrak{c}} \mathfrak{c} =_{\mathfrak{c}} \mathcal{P}, \quad (10.26)$$

for example $I = \mathfrak{c}$, and bijections

$$\alpha \mapsto x_\alpha \in \mathcal{N}, \quad \alpha \mapsto P_\alpha \in \mathcal{P} \quad (\alpha \in I)$$

which witness the equinumerosities (10.26). Fix also a best wellordering $\leq$ of I . We will define by transfinite recursion on $(I, \leq)$ pointsets

$$A_\alpha, B_\alpha \subset \mathcal{N} \quad (\alpha \in I),$$

so that the following conditions hold.

1. $A_\alpha \cap B_\alpha = \emptyset$, for each $\alpha \in I$.
2. $\alpha \leq \beta \implies A_\alpha \subseteq A_\beta, B_\alpha \subseteq B_\beta$.
3. $|A_\alpha| =_{\mathfrak{c}} |B_\alpha| =_{\mathfrak{c}} |\mathbf{seg}(\alpha)|$.
4. For each $\alpha \in I$, $B_{S\alpha} \cap P_\alpha \neq \emptyset$, where S is the successor function in the well ordered set $(I, \leq)$.

Lemma 2. If A_α, B_α ($\alpha \in I$) satisfy (1) - (4), then the union

$$A = \bigcup_{\alpha \in I} A_\alpha$$

has no non-empty, perfect subset, but $|A| =_c \mathfrak{c}$.

Proof. $A =_c I =_c \mathfrak{c}$ follows immediately from (3). If $P \neq \emptyset$ is perfect, then $P = P_\alpha$ for some $\alpha \in I$, hence there exists some $x \in P_\alpha \cap B_{S_\alpha}$ and then $x \notin A$, by (4) and (1).

The definition of A_α, B_α is practically forced on us by conditions (1) - (4). We outline the argument for the proof of (1) - (4) by transfinite induction together with the clauses of the transfinite recursion definition, pedantically it should be separated out and explained after the definition is completed.

(a) At the minimum 0 of I , set $A_0 = B_0 = \emptyset$.

(b) If λ is a limit point of I , set

$$A_\lambda = \bigcup_{\alpha < \lambda} A_\alpha, \quad B_\lambda = \bigcup_{\alpha < \lambda} B_\alpha.$$

Conditions (1) and (2) hold trivially, (4) is not involved in this case, and (3) is verified by applying 9.17 both ways. For example,

$$\alpha < \lambda \implies |A_\alpha| =_c |\mathbf{seg}(\alpha)| \leq_c |\mathbf{seg}(\lambda)|$$

by the induction hypothesis, hence by 9.17

$$|A_\lambda| =_c \left| \bigcup_{\alpha < \lambda} A_\alpha \right| \leq_c \sum_{\alpha < \lambda} |A_\alpha| |\mathbf{seg}(\lambda)| \leq_c |\mathbf{seg}(\lambda)|.$$

(c) Suppose $\beta = S\alpha$ is a successor point in I . By the induction hypothesis, each of A_α and B_α are equinumerous with $\mathbf{seg}(\alpha)$ and $\mathbf{seg}(\alpha) <_c I$, because $\leq$ is a best wellordering. Thus, $|A_\alpha|, |B_\alpha|$ are both smaller than $\mathfrak{c}$, hence $|A_\alpha \cup B_\alpha| <_c \mathfrak{c}$, and we can find in the non-empty, perfect set $P_\alpha =_c \mathcal{N}$ distinct points

$$x, y \in P_\alpha \setminus (A_\alpha \cup B_\alpha);$$

we set

$$A_\beta = A_\alpha \cup \{x\}, \quad B_\beta = B_\alpha \cup \{y\}.$$

Conditions (1) and (2) are trivial, (4) is insured by the definition of $B_\beta = B_{S\alpha}$, and (3) is immediate from the induction hypothesis, since each of $A_\beta, B_\beta, \mathbf{seg}(\beta)$ has just one more point than its respective predecessor $A_\alpha, B_\alpha, \mathbf{seg}(\alpha)$. $\dashv$

The construction obviously proves more than what is claimed in the theorem: $|A| =_c \mathfrak{c}$ and both A and its complement ${}_c A$ intersect every non-empty, perfect set. We leave for the problems some additional variations

which make it even more obvious that the program of proving the Continuum Hypothesis by using the Cantor-Bendixson Theorem is hopeless.

Actually, it is not only this program for settling the Continuum Problem which fails: every attempt to prove or disprove **CH** from the axioms of **ZAC** is doomed, by the following two central independence results.

10.27. Consistency of GCH, the Generalized Continuum Hypothesis (Gödel, 1939). *The model L of constructible sets satisfies the Generalized Continuum Hypothesis **GCH**, (3.2), so, in particular, the Continuum Hypothesis cannot be refuted in **ZAC**.*

10.28. Independence of the Continuum Hypothesis CH (Cohen, 1963). *There is a model of **ZAC** in which the Continuum Hypothesis fails, hence **CH** cannot be proved in **ZAC**. Cohen's forcing model can be modified in many ways to manipulate the cardinalities of pointsets and subsets of larger powersets.*

10.29. What does the independence of CH mean? Both the Gödel and the Cohen methods of proof are very robust and they have been adapted to show that the Continuum Problem cannot be settled on the basis of many reasonable and plausible strengthenings of **ZAC** by additional axioms. The same is true of the Axiom of Choice, of course, or the Axiom of Infinity for that matter, but it is clear that these propositions express new, fundamental set theoretic principles which are most likely true but cannot (and, in fact, cannot be expected to) be proved from simpler axioms by logic alone. The Continuum Hypothesis has the look of a technical, mathematical problem which should be settled definitively by a proof, but we seem to lack the insight needed to divine the necessary axioms.

Much has been made of this independence of **CH** (and many more assertions about sets) from variants of the known axioms of set theory, and some have used it to argue against any objective reality behind the "formal," axiomatic results of the subject. Using the method of *arithmetization* introduced by Gödel, however, we can translate questions about the *existence of proofs* into precise, technical conjectures about integers: since there exist such conjectures⁵ which (like **CH**) can be shown to be undecidable in the known, plausible axiomatic theories, are we then forced to deny objective reality to the natural numbers? In fact, it is not possible to discuss such

⁵The type of statements we have in mind here are of the form "if **ZFC** is consistent, then so is T ," where T is some strong extension of **ZFC** which, in fact, implies the consistency of **ZFC**. Gödel's *Second Incompleteness Theorem* implies that statements of this type are independent of **ZFC** (unless **ZFC** proves its own inconsistency), and there are many of them about whose truth there is genuine controversy.

problems intelligently without reference to notions and results of Mathematical Logic which are beyond the scope of these Notes, and we will resist the temptation.

Incidentally, there are scores of interesting propositions about sets which cannot be settled on the basis of **ZDC** or **ZAC**, **CH** is only the most interesting of them. We mention here just three more independence results of this type, because they are relevant to the Perfect Set Theorem.

10.30. (Gödel, 1939) *In the model L of constructible sets, there exists an uncountable, co-analytic set which has no proper perfect subset.* This means that we cannot improve the Perfect Set Theorem **10.18** in **ZAC** to show that every co-analytic pointset has property **P**.

10.31. (Solovay, 1970) *There is a model of **ZAC** in which every “definable” pointset has property **P**.* We will not attempt to define “definable,” but analytic complements are definable.

10.32. (Solovay, 1970) *There is a model of **ZDC** in which every pointset has property **P**.*

Solovay’s models are constructed by Cohen’s forcing method, but like Gödel’s L , they have many more canonical properties which yield numerous unprovability results. The first Solovay model witnesses (with **10.30**) that the property **P** for analytic complements cannot be proved or refuted in **ZAC**. The second Solovay model shows that **ZDC** cannot prove the existence of an uncountable pointset with no non-empty, perfect subset; **DC** is not a sufficiently strong choice principle to effect the construction.

Problems

x10.1. Prove that the decomposition (10.11) of a closed pointset F into a perfect set P and a countable set S determines uniquely P and S .

x10.2. Give an example of a closed pointset $F \subseteq \mathcal{N}$ and a continuous $f : \mathcal{N} \rightarrow \mathcal{N}$, such that the image $f[F]$ is not closed.

x10.3. Prove that every open pointset is an $\mathcal{F}_\sigma$ and every closed pointset is a $\mathcal{G}_\delta$. The definitions are reviewed in Footnote 3.

***x10.4.** Prove that the inverse image $g^{-1}[A]$ of an analytic pointset A by a continuous function $g : \mathcal{N} \rightarrow \mathcal{N}$ is analytic. **HINT:** Aim for an equivalence of the form

$$y \in g^{-1}[A] \iff (\exists x)[y = f(\rho_1(x)) = g(\rho_2(x))]$$

where f is continuous and ρ_n are defined by (10.22), and then use **10.21**.

***x10.5.** Prove that

$$\mathcal{N} = \bigcup_{i \in \mathbb{N}} A_i \implies (\exists i)[A_i =_c \mathcal{N}],$$

i.e. $\mathcal{N}$ is not the union of a countable sequence of pointsets of smaller cardinality. **HINT:** This follows easily from König's Theorem **9.21**, but the relevant special case does not need the full Axiom of Choice.

x10.6. (AC) Prove that for every $\kappa \leq_c \mathfrak{c}$, there exists a pointset A with $|A| =_c \kappa$ which contains no non-empty, perfect subset.

x10.7. (AC) Prove that there exists an uncountable pointset A such that neither A nor its complement contain an uncountable Borel set.

x10.8. Prove that there are $\mathfrak{c}$ -many analytic and Borel pointsets, $|\mathcal{A}| =_c |\mathcal{B}| =_c \mathfrak{c}$.

10.33. Definition. A function $f : X \rightarrow Y$ from one topological space into another is **Borel measurable** if the inverse image $f^{-1}[G]$ of every open subset of Y is a Borel subset of X .

x10.9. The composition $gf : X \rightarrow Z$ of two Borel measurable functions $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ is Borel measurable.

x10.10. The inverse image $f^{-1}[A]$ of a Borel set $A \subseteq Y$ by a Borel measurable function $f : X \rightarrow Y$ is a Borel subset of X .

10.34. Definition. Two topological spaces X, Y are **Borel isomorphic** if there exists a bijection $f : X \xrightarrow{\sim} Y$ such that both f and its inverse $f^{-1} : Y \xrightarrow{\sim} X$ are Borel measurable. Borel isomorphic spaces have the same measure-theoretic structure and for all practical purposes can be “identified” in measure theory.

***x10.11.** Suppose $f : X \xrightarrow{\sim} Y$ and $g : Y \xrightarrow{\sim} X$ are Borel measurable injections between topological spaces, with the following additional property: there exists Borel measurable functions $f_1 : Y \rightarrow X$ and $g_1 : X \rightarrow Y$ which are inverses of f and g in the sense that

$$\begin{aligned} f_1 f(x) &= x & (x \in X), \\ g_1 g(y) &= y & (y \in Y). \end{aligned}$$

Prove that X and Y are Borel isomorphic. **HINT:** Use the proof of the Schröder-Bernstein Theorem **2.24**.

x10.12. Consider the Cantorset $\mathcal{C}$ as a topological subspace of $\mathcal{N}$ in the obvious way, the open sets being unions of neighborhoods of the form

$$\mathcal{N}_u = \{x \in \mathcal{C} \mid u \sqsubseteq x\} \quad (u \in \{0, 1\}^*).$$

Prove that $\mathcal{C}$ and $\mathcal{N}$ are Borel isomorphic.

In the remaining problems we explore the connection of the specific, combinatorial notions we studied in Baire space with their general, topological versions.

10.35. Definition. A point x is a **limit point** of a set A in a topological space X if every open set which contains x contains also some point of A other than x ,

$$(\forall G)[G \text{ open and } x \in G \implies (\exists y \in A \cap G)[x \neq y]].$$

A limit point of A may or may not be a member of A . A point of A which is not a limit point of A is **isolated in A** .

x10.13. Determine the limit points and the isolated points of the pointset

$$B = \{x \in \mathcal{N} \mid x(0) = 1 \vee (\forall n)[x(n) = 2] \vee (\exists n)[x(n) = 3]\}.$$

x10.14. Prove that x is a limit point of A if and only if every open set containing x contains infinitely many points of A .

x10.15. Prove that a set is closed in a topological space X if and only if it contains all its limit points.

x10.16. Prove that a pointset P is perfect if and only if it is closed and has no isolated points, i.e. every point of P is a limit point of P . This equivalence identifies the specific definition of perfect pointsets we adopted with the classical, topological definition.

10.36. Definition. A sequence $(n \mapsto x_n)$ of points in a topological space X **converges** to a point x or has x as its **limit** if every open set containing x contains all but finitely many of the terms of the sequence,

$$\lim_n x_n = x \iff_{\text{df}} (\forall G \text{ open}, x \in G)(\exists n \in \mathbb{N})(\forall i \geq n)[x_i \in G].$$

x10.17. Prove that a point x is a limit point of a pointset A if and only if $x = \lim_n x_n$ is the limit of some sequence $(n \mapsto x_n \in A)$ of points in A . Which choice principle did you use, if any?

x10.18. Prove that a function $f : \mathcal{N} \rightarrow \mathcal{N}$ is continuous if and only if

$$f(\lim_n x_n) = \lim_n f(x_n),$$

whenever $\lim_n x_n$ exists. Which choice principle did you use, if any?

x10.19. A topological space X is **Hausdorff** if for any two points $x \neq y$, there exist disjoint open sets $G \cap H = \emptyset$ such that $x \in G$, $y \in H$. Prove that if $f, g : X \rightarrow Y$ are continuous and Y is Hausdorff, then the set $\{x \in X \mid f(x) = g(x)\}$ is closed in X .

10.37. Definition. An **open covering** of a set K in a topological space X is any family $\mathcal{G}$ of open sets whose union includes K , $K \subseteq \bigcup \mathcal{G}$. A set K is **compact** in X if every open covering of K includes a finite subcovering, i.e. for every family $\mathcal{G}$ of open sets,

$$K \subseteq \bigcup \mathcal{G} \implies (\exists G_0, \dots, G_n \in \mathcal{G})[K \subseteq \bigcup_{n \leq i} G_i].$$

***x10.20.** Prove that a pointset is compact by Definition **10.16** if and only if it is compact by Definition **10.37**. **HINT:** You will need König's Lemma.

x10.21. Prove that for any two topological spaces X, Y , any continuous function $f : X \rightarrow Y$ and any compact set $K \subseteq X$, the image $f[K]$ is compact in Y .

Chapter 11

REPLACEMENT AND OTHER AXIOMS

We have just about reached one of the goals we set in Chapter 4, which was to prove all the “naive” results of Chapter 2 from the axioms of Zermelo. Only a couple of minor points remain, but they are significant: they will reveal that Zermelo’s axioms are not sufficient and must be supplemented by stronger principles of set construction. Here we will formulate and add to the axiomatic theory **ZDC** the **AXIOM OF REPLACEMENT** discovered in the early 1920’s, a principle of set construction no less plausible than any of the constructive axioms **(I) - (VI)** but powerful in its consequences. We will also introduce and discuss some additional principles which are often included in axiomatizations of set theory. Using only a weak consequence of Replacement, we will construct the **LEAST ZERMELO UNIVERSE** $\mathcal{Z}$, a remarkably simple set which contains the natural numbers, Baire space, the real numbers and all the significant objects of study of classical mathematics. Everything we have proved so far can be interpreted as if $\mathcal{Z}$ comprised the entire universe of mathematical objects, yet $\mathcal{Z}$ is just a set—and a fairly small, easy to comprehend set, at that! Our main purpose in this chapter is to understand the Axiom of Replacement by investigating its simplest and most direct consequences. The real power of this remarkable proposition will become apparent in the next chapter.

According to (2) of **2.16**, if A is a countable set and for each $n \geq 2$,

$$A^n = \underbrace{A \times \cdots \times A}_{n \text{ times}},$$

then the union $\bigcup_{n=2}^{\infty} A^n$ is also countable. The obvious way to prove this from the axioms is to define first the sets A^n by the recursion

$$\begin{aligned} f(0) &= A \times A, \\ f(n+1) &= f(n) \times A, \end{aligned} \tag{11.1}$$

so that $f(n) = A^{n+2}$ and

$$\bigcup_{n=0}^{\infty} A^{n+2} = \bigcup f[N]; \tag{11.2}$$

Cantor’s basic **2.10** implies first (by induction) that each $f(n) = A^{n+2}$ is countable, and then that their union $\bigcup_{n=0}^{\infty} A^{n+2}$ must also be countable. Is

there an error? Certainly not in the proof by induction, which is no different than many others like it. There is a problem, however, with the recursive definition (11.1) which cannot be justified by the Recursion Theorem 5.6 as it stands. To apply 5.6 we need a set E , a function $h : E \rightarrow E$ on E and some $a \in E$, which then determine a unique $f : N \rightarrow E$ satisfying

$$\begin{aligned} f(0) &= a, \\ f(n+1) &= h(f(n)). \end{aligned} \tag{11.3}$$

In the case at hand there is no obvious E which contains A and all its products A^n , and instead of a function h , we have the operation

$$h(X) =_{\text{df}} X \times A, \tag{11.4}$$

which associates with each set X its product $X \times A$ with the given set A . To justify definition (11.1), we need a recursion theorem which validates recursive definitions of the form (11.3), for every object a and (unary) definite operation h . It looks quite innocuous, only a mild generalization of the Recursion Theorem—and it is just that—but in fact such a result cannot be established rigorously on the basis of the Zermelo axioms.

11.1. (VIII) Replacement Axiom. *For each set A and each unary definite operation H , the image*

$$H[A] =_{\text{df}} \{H(x) \mid x \in A\}$$

of A by H is a set. As a construction principle for sets, the Replacement Axiom is almost obvious, as plausible on intuitive grounds as the Separation Axiom. If we already understand A as a completed totality and H associates in a definite and unambiguous manner an object with each $x \in A$, then we can “construct” the image $H[A]$ by “replacing” each $x \in A$ by the corresponding $H(x)$.

11.2. Axiomatics. *The axiomatic system **ZFDC** of Zermelo-Fraenkel set theory with Dependent Choices comprises the axioms of **ZDC** and the Replacement Axiom 11.1, symbolically*

$$\mathbf{ZFDC} = \mathbf{ZDC} + \text{Replacement} = (\mathbf{I}) - (\mathbf{VIII}).$$

*The corresponding system with the full Axiom of Choice is **ZFAC**, symbolically*

$$\mathbf{ZFAC} = \mathbf{ZAC} + \text{Replacement} = (\mathbf{I}) - (\mathbf{VIII}) + \mathbf{AC}.$$

From now on we will use all the axioms of **ZFDC** without explicit mention

and we will continue to annotate by the mark **(AC)** the results of **ZFAC**, whose proof requires the full Axiom of Choice.¹

Mostly we have used simple definite operations up until now, those directly supplied by the axioms like $\mathcal{P}(A)$ and $\bigcup \mathcal{E}$ and explicit combinations of them, e.g. the Kuratowski pair $(x, y) =_{\text{df}} \{\{x\}, \{x, y\}\}$. Once we assume the Axiom of Replacement, however, definite operations come into center stage and we will need to deal with some which are not so simply defined. We describe in the next, trivial Proposition the basic method of definition we will use, primarily to point attention to it.

11.3. Proposition. *Suppose C and P are definite conditions of n and $n + 1$ arguments, respectively, assume that*

$$(\forall \vec{x})[C(\vec{x}) \implies (\exists! w)P(\vec{x}, w)], \quad (11.5)$$

and let

$$F(\vec{x}) =_{\text{df}} \begin{cases} \text{the unique } w \text{ such that } P(\vec{x}, w), & \text{if } C(\vec{x}), \\ \emptyset, & \text{otherwise.} \end{cases} \quad (11.6)$$

The n -ary operation F is definite. ⊣

In practice, we will appeal to this observation by setting

$$F(\vec{x}) =_{\text{df}} \text{the unique } w \text{ such that } P(\vec{x}, w) \quad (C(\vec{x})), \quad (11.7)$$

after we verify (11.5), without specifying the irrelevant value of F outside the domain we care about. The Axiom of Replacement often comes into the proof of (11.5).

11.4. Exercise. *For each unary definite operation F , the operation*

$$G(X) =_{\text{df}} F[X] = \{F(x) \mid x \in X\} \quad (\text{Set}(X))$$

is also definite.

The next fundamental consequence of the Replacement Axiom generalizes the Transfinite Recursion Theorem in two ways: by allowing a definite operation instead of just a function in the statement, and by replacing the given well ordered set by an arbitrary grounded graph. The second generalization does not require the Replacement Axiom, Problem *x8.11.

¹It would be historically more accurate to honor the great mathematician and logician Skolem for the discovery of the Replacement Axiom, but the use of Fraenkel's name and the letter **F** in acronyms to signify inclusion of the Replacement Axiom has been sanctified by long-term usage and it is not realistic to try and change it. Fraenkel considered propositions similar to the Replacement Axiom, as in fact Cantor had done, much earlier.

11.5. Grounded Recursion Theorem. *For each grounded graph G with edge relation $\rightarrow$ and each unary definite operation H , there exists exactly one function $f : G \rightarrow f[G]$ which satisfies the identity*

$$f(x) = H(f \upharpoonright \{y \in G \mid x \rightarrow y\}).$$

Proof. As in the proof of 7.24, we first show a lemma which gives us a set of approximations of the required function. Instead of the initial segments of G (which do not make much sense for an arbitrary graph), these approximations are defined here on downward closed subsets of G . Recall the definition of the transitive closure of a graph $\Rightarrow_G$ given in 6.33; we will skip all the subscripts in what follows, since only the single graph G is involved in the argument, and we will also use the inverse arrows,

$$u \leftarrow t \iff_{\text{df}} t \rightarrow u \iff u \text{ is immediately below } t, \quad (11.8)$$

$$x \Leftarrow t \iff_{\text{df}} t \Rightarrow x \iff x \text{ is (on some path) below } t. \quad (11.9)$$

Lemma. For each node $t \in G$, there exists exactly one function σ with domain the set $\{x \in G \mid x \Leftarrow t\}$ which satisfies the identity

$$\sigma(x) = H(\sigma \upharpoonright \{y \in G \mid y \Leftarrow x\}) \quad (x \Leftarrow t). \quad (11.10)$$

Proof. Suppose, towards a contradiction, that t is a minimal node of G where the Lemma fails. Thus, for each $u \leftarrow t$, there is exactly one function σ_u such that

$$\sigma_u(x) = H(\sigma_u \upharpoonright \{y \in G \mid y \Leftarrow x\}) \quad (x \Leftarrow u). \quad (11.11)$$

First we notice that

$$x \Leftarrow u \leftarrow t \ \& \ x \Leftarrow v \leftarrow t \implies \sigma_u(x) = \sigma_v(x); \quad (11.12)$$

because if x were minimal in G where (11.12) failed, then

$$\begin{aligned} \sigma_u(x) &= H(\sigma_u \upharpoonright \{y \in G \mid y \Leftarrow x\}) \quad \text{by (11.11),} \\ &= H(\sigma_v \upharpoonright \{y \in G \mid y \Leftarrow x\}) \quad \text{by the choice of } x, \\ &= \sigma_v(x) \quad \text{by (11.11) for } \sigma_v. \end{aligned}$$

The operation $u \mapsto \sigma_u$ which assigns this σ_u to each $u \leftarrow t$ is definite, so by the Axiom of Replacement its image is a set and we can set

$$\sigma_1 =_{\text{df}} \bigcup \{\sigma_u \mid u \leftarrow t\};$$

this σ_1 is a function by (11.12), and by the definition,

$$\sigma_1(x) \downarrow \iff (\exists u)[x \Leftarrow u \leftarrow t].$$

By another application of the Replacement Axiom,

$$\sigma_2 =_{\text{df}} \{(v, H(\sigma_v \upharpoonright \{x \mid x \leftarrow v\})) \mid v \leftarrow t \ \& \ \neg(\exists u)[v \leftarrow u \leftarrow t]\}$$

is also a set, and by its definition it is a function with domain disjoint from that of σ_1 . Thus

$$\sigma =_{\text{df}} \sigma_1 \cup \sigma_2$$

is a function, and

$$\begin{aligned} \sigma(x) \downarrow &\iff (\exists u)[t \rightarrow u \ \& \ u \Rightarrow x] \vee t \rightarrow u \\ &\iff t \Rightarrow x \end{aligned} \quad \text{by 6.34.}$$

Moreover, σ satisfies (11.10) because σ_1 and σ_2 do. Finally, the same argument by which we proved (11.12) shows that no more than one σ with domain $\{x \in G \mid x \leftarrow t\}$ can satisfy (11.10), and that completes the proof of the Lemma.

To prove the Theorem, we apply the Lemma as in 7.24 to “the successor graph”

$$\begin{aligned} \text{Succ}(G) &=_{\text{df}} G \cup \{t\}, \\ x \rightarrow_{\text{Succ}(G)} y &\iff x \rightarrow y \vee [x = t \ \& \ y \in G], \end{aligned}$$

which has just one more node than G , at the top.

11.6. Corollary. (1) *For each well ordered set U and each unary definite operation H , there exists exactly one function $f : U \rightarrow f[U]$ which satisfies the identity*

$$f(x) = H(f \upharpoonright \text{seg}(x)) \quad (x \in U). \quad (11.13)$$

(2) *For each object a and each definite unary operation F , there exists a unique sequence $(n \mapsto a_n)$ which satisfies the identities*

$$a_0 = a, \quad a_{n+1} = F(a_n) \quad (n \in N). \quad (11.14)$$

We call $(n \mapsto a_n)$ the **orbit** of a under F .

Proof. For (1) we apply 11.5 to the graph $(\text{Field}(U), >_U)$, and for (2) to the graph $(N, \rightarrow)$, where

$$n \rightarrow m \iff n = m + 1. \quad \dashv$$

11.7. Exercise. Which definite operation H do we use to prove (2) of the Corollary?

The orbit of a set A by the unionset operation reveals the hidden structure of A under the membership relation by exposing the members of A , the members of the members of A , the members of those, etc. ad infinitum.

11.8. Definition. A class M is **transitive** if $\bigcup M \subseteq M$, equivalently $(\forall x \in M)(\forall t \in x)[t \in M]$, or just $x \in M \implies x \subseteq M$.

11.9. Exercise. The sets $\emptyset, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, the set

$$N_0 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\} \quad (11.15)$$

postulated by the Axiom of Infinity and every class whose members are all atoms are transitive.

11.10. Transitive Closure Theorem. Every set A is a member of some transitive set M , in fact, there is a least (under $\subseteq$) transitive set $M = TC(A)$ such that $A \in TC(A)$. We call $TC(A)$ the **transitive closure** of A .

Proof. By (2) of 11.6, there is a unique sequence $n \mapsto TC_n(A)$ which satisfies the identities

$$\begin{aligned} TC_0(A) &= \{A\}, \\ TC_{n+1}(A) &= \bigcup TC_n(A), \end{aligned} \quad (11.16)$$

and we set

$$TC(A) =_{\text{df}} \bigcup_{n=0}^{\infty} TC_n(A). \quad (11.17)$$

Clearly $A \in TC(A)$ and $TC(A)$ is transitive, because

$$u \in TC_n(A) \implies u \subseteq \bigcup TC_n(A) = TC_{n+1}(A).$$

If M is transitive and $A \in M$, then $TC_0(A) = \{A\} \subseteq M$, and by induction

$$TC_n(A) \subseteq M \implies TC_{n+1}(A) = \bigcup TC_n(A) \subseteq \bigcup M \subseteq M,$$

so that in the end $TC(A) = \bigcup_n TC_n(A) \subseteq M$. ⊣

11.11. Exercise. If A is transitive, then $TC(A) = A \cup \{A\}$.

To understand better the remark about “revealing the hidden $\in$ -structure” of A , consider the following natural concepts.

11.12. Definition. A set A is **hereditarily free of atoms** or **pure** if it belongs to some transitive set which contains no atoms; equivalently, if $TC(A)$ contains no atoms. A set A is **hereditarily finite** if it belongs to some transitive, finite set; equivalently, if $TC(A)$ is finite. A set A is **hereditarily countable** if it belongs to some transitive, countable set; equivalently, if $TC(A)$ is countable.

The point of the definitions is that $\{\{a\}\}$ is a set but not a pure set if a is an atom, because we need a to construct it; $\{N\}$ is finite but not hereditarily finite because we need all the natural numbers to construct it; $\{\mathcal{N}\}$ is countable but not hereditarily countable because we need to “collect into a whole” an uncountable collection of objects in $\mathcal{N}$ before we can construct the singleton $\{\mathcal{N}\}$ by one final, trivial act of collection. Put another way, $\{\mathcal{N}\}$ is not hereditarily countable because “its concept involves” an uncountable infinity of objects, the members of its sole member $\mathcal{N}$.

11.13. Exercise. *A transitive class is pure exactly when it has no atoms. Consequently, the Principle of Purity 3.24 is equivalent to the assertion that every set is pure.*

11.14. Exercise. *A transitive set is hereditarily finite if it is finite, and hereditarily countable if it is countable.*

Next we consider the closure of a set under both the unionset and powerset operations.

11.15. Basic Closure Lemma. *For each set I and each natural number n , let $M_n = M_n(I)$ be the set defined by the recursion*

$$M_0 = I, \quad M_{n+1} = M_n \cup \bigcup M_n \cup \mathcal{P}(M_n). \quad (11.18)$$

The basic closure of I is the union

$$M = M(I) =_{\text{df}} \bigcup_{n=0}^{\infty} M_n(I), \quad (11.19)$$

and it has the following properties.

(1) *M is a transitive set which contains $\emptyset$ and I , it is closed under the pairing $\{x, y\}$, unionset $\bigcup \mathcal{E}$ and powerset $\mathcal{P}(A)$ operations and it contains every subset of each of its elements.*

(2) *M is the least (under $\subseteq$) transitive set which contains I and is closed under $\{x, y\}$, $\bigcup \mathcal{E}$ and $\mathcal{P}(X)$.*

(3) *If I is pure and transitive, then each M_n is a pure, transitive set and satisfies*

$$M_{n+1} = \mathcal{P}(M_n). \quad (11.20)$$

As a consequence, M is a pure, transitive set.

Proof. (1) By the definition, $\emptyset, I \in M_1 \subseteq M$. If $x, y \in M$, then from the obvious $M_n \subseteq M_{n+1}$, there exists some m such that $\{x, y\} \subseteq M_m$, so $\{x, y\} \in M_{m+1}$. The key inclusion for the remaining claims is

$$x \in M_n \implies x \subseteq \bigcup M_n \subseteq M_{n+1} \subseteq M,$$

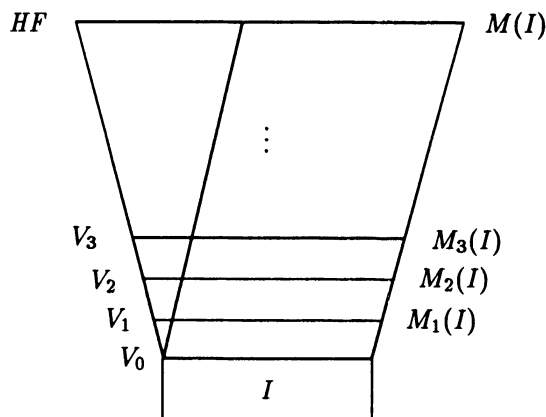


Figure 11.1. Logarithmic² renditions of $M(I)$ and HF .

which implies immediately that M is transitive. It also implies

$$x \in M_n \implies \bigcup x \subseteq \bigcup M_{n+1} \subseteq M_{n+2},$$

so $x \in M_n \implies \bigcup x \in M_{n+3} \subseteq M$ and M is closed under $\bigcup x$. The same argument shows that M is closed under $\mathcal{P}(X)$ and the last assertion follows by this closure and transitivity.

(2) Closure of any M' under $\{x, y\}$ and $\bigcup \mathcal{E}$ implies closure under $A \cup B = \bigcup \{A, B\}$, and the additional closure under $\mathcal{P}(X)$ implies by a simple induction that $M_n \in M'$ for each n , so $M \subseteq M'$ by the transitivity of M' .

(3) If I is transitive with no atoms, then every M_n is transitive and has no atoms by a trivial induction on n . This implies that M is a transitive set with no atoms and hence pure, but also that $M_n \cup \bigcup M_n \subseteq \mathcal{P}(M_n)$, so that $M_{n+1} = \mathcal{P}(M_n)$. $\dashv$

11.16. Exercise. In connection with the proof of (3), is not the inclusion $X \subseteq \mathcal{P}(X)$ true for every transitive set X ?

11.17. Exercise. If $I \subseteq J$, then for each n , $M_n(I) \subseteq M_n(J)$, and hence $M(I) \subseteq M(J)$.

11.18. The Hereditarily Finite Sets. The least basic closure is that of the empty set, $HF =_{\text{df}} M(\emptyset) \subseteq M(I)$, for every I . In the classical notation, $M_n(\emptyset) = V_n$, so that the V_n 's and their union are determined by

²Picturing universes of sets by cones like this is traditional but misleading; the successive powersets grow hyperexponentially in size, so it would be more accurate to draw a cone with curved, hyperexponential sides.

the identities

$$V_0 = \emptyset, \quad V_{n+1} = \mathcal{P}(V_n), \quad HF =_{\text{df}} \bigcup_{n=0}^{\infty} V_n = M(\emptyset). \quad (11.21)$$

For example, $\emptyset \in V_1$, $\{\emptyset\} \in V_2$ and $\{\{\emptyset\}, \{\{\emptyset\}\}\} \in V_4$! The set HF is pure and transitive, each V_n is finite by an easy induction, so every set in HF is pure and hereditarily finite, and HF itself is countable. These are the sets which can be constructed “from nothing” (literally, the empty set) by iterating any finite number of times the operation of collecting into a whole (putting between braces) some of the objects already constructed.

The closure properties of $M(I)$ itemized in (1) of **11.15** are precisely those required of the universe $\mathcal{W}$ by axioms **(II)** - **(V)**, as we discussed them in **3.25**, and if the set N_0 of (11.15) demanded by the Axiom of Infinity **(VI)** is a subset of I , we also have $N_0 \in M(I)$. Notice also that since $M(I)$ is transitive, for $A, B \in M(I)$,

$$A \neq B \implies (\exists t \in M(I))[t \in (A \setminus B) \cup (B \setminus A)], \quad (11.22)$$

which says of $M(I)$ what the Axiom of Extensionality demands of $\mathcal{W}$ by (3.10). This suggests that if we take “object” to mean “member of $M(I)$,” for any $I \supseteq N_0$, then we can reinterpret every proof from the axioms **(II)** - **(V)** as an argument about the members of $M(I)$ instead of all objects, in the end proving a theorem about $M(I)$ instead of $\mathcal{W}$. It is an important idea, worth abstraction and a name.

11.19. Definition. *A transitive class M is a Zermelo universe if it is closed under the pairing $\{x, y\}$, unionset $\bigcup \mathcal{E}$ and powerset $\mathcal{P}(X)$ operations, and contains the set N_0 defined by (11.15). The least Zermelo universe is $\mathcal{Z} = M(N_0)$, determined by the identities*

$$\mathcal{Z}_0 = N_0, \quad \mathcal{Z}_{n+1} = \mathcal{P}(\mathcal{Z}_n), \quad \mathcal{Z} = \bigcup_{n=0}^{\infty} \mathcal{Z}_n. \quad (11.23)$$

11.20. Exercise. *The class $\mathcal{W}$ of all objects is a Zermelo universe. Every Zermelo universe contains the empty set as well as every subset of each of its members.*

A Zermelo universe M is a *model* of the axioms **(I)** - **(VI)**, and a very special model at that, since it interprets standardly the basic relations of membership and sethood—it only restricts the domain of objects in which we interpret propositions. The claim that logical consequences of **(I)** - **(VI)** are true in every Zermelo universe is called a *metatheorem*, a theorem about theorems. To make general results of this type completely precise and prove them rigorously requires concepts from *Mathematical Logic*. In specific instances, however, lemma by lemma and proposition by proposition, it is quite simple to see what the specific consequence of the

axioms means for an arbitrary domain of objects and to verify it in every Zermelo universe: this is because, in fact, we have been using the axioms as closure properties of the universe, about which we have assumed nothing more but that it satisfies them.

11.21. Proposition. *Every Zermelo universe M is closed under the Kuratowski pair operation (x, y) defined in (4.1), as well as the Cartesian product $A \times B$, function space $(A \rightarrow B)$, and partial function space $(A \rightharpoonup B)$ operations, provided these are defined using the Kuratowski pair. In addition, if $A \in M$ and $\sim$ is an equivalence relation on A , then $\sim$ and the quotient $\llbracket A/\sim \rrbracket$ defined in 4.14 are also in M .*

Proof. The Kuratowski pair $(x, y) = \{\{x\}, \{x, y\}\}$ of any two members of M is obtained by taking unordered pairs twice, so it is certainly in M . If $A, B \in M$, then $A \cup B = \bigcup \{A, B\}$, and by the proof of 4.3, $A \times B \subseteq \mathcal{P}(\mathcal{P}(A \cup B)) \in M$, so $A \times B \in M$. The rest are proved similarly. $\dashv$

11.22. Exercise. *Every Zermelo universe M contains a system of natural numbers as defined in 5.1.*

11.23. Proposition. (1) *The Axiom of Dependent Choices is true in every Zermelo universe M , in the following sense: if $a \in A \in M$, $P \subseteq A \times A$, $P \in M$ and $N \in M$ is a system of natural numbers in M , then*

$$a \in A \ \& \ (\forall x \in A)(\exists y \in A)xPy \\ \implies (\exists f : N \rightarrow A)[f \in M \ \& \ f(0) = a \ \& \ (\forall n \in N)f(n)Pf(n+1)].$$

(2) **(AC)** *The Axiom of Choice is true in every Zermelo universe M , in the following sense, following 8.4: for every family $\mathcal{E} \in M$ of non-empty and pairwise disjoint sets, then there exists some set $S \in M$ which is a choice set for $\mathcal{E}$, i.e.*

$$S \subseteq \bigcup \mathcal{E}, \quad (\forall X \in \mathcal{E})(\exists u)[S \cap X = \{u\}].$$

Proof. (1) The hypothesis of the implication to be proved implies by **DC** that there exists some function $f : N \rightarrow A$ such that $f(0) = a$ and for every $n \in N$, $f(n)Pf(n+1)$. Since $(N \rightarrow A) \in M$, we also have $f \in M$ by transitivity.

Part (2) is proved similarly. $\dashv$

Although we chose specific versions of the choice principles to simplify these arguments, their numerous equivalents are also true in every Zermelo universe M . This can be verified directly, or by observing that the equivalence proofs we have given can be “carried out within M .”

11.24. Exercise. (AC) If M is any Zermelo universe, $A, B \in M$, and P is any binary definite condition, then

$$(\forall x \in A)(\exists y \in B)P(x, y) \implies (\exists f \in M)[f : A \rightarrow B \ \& \ (\forall x \in A)P(x, f(x))].$$

11.25. The least Zermelo universe. Let us concentrate on the least Zermelo universe $\mathcal{Z}$, to focus the argument. It is a pure, transitive set, constructed by starting with the simple set N_0 and iterating the powerset operation infinitely many times, much as we construct the natural numbers starting with 0 and iterating infinitely many times the successor operation. We can think of the sets in $\mathcal{Z}$ as precisely those objects whose existence is guaranteed by the axioms (I) - (VI). The natural interpretations of DC and AC are also true in $\mathcal{Z}$, the latter under the assumption that AC is true in $\mathcal{W}$. Using the closure properties of Zermelo universes already established and looking back at Chapters 5, 6 and 10 and ahead at Appendix A, we can verify that $\mathcal{Z}$ contains not only the specific system of natural numbers N we constructed in Chapter 5, but also the Baire space $\mathcal{N}$ defined from this N and the specific systems of rational and real numbers constructed in Appendix A. By the uniqueness results, any one of these systems is as good as any other, so we can say that $\mathcal{Z}$ contains the integers, Baire space, the rationals and the reals.

Combining these remarks with some knowledge of classical mathematics, it is not hard to give a convincing argument that *all the objects studied in classical algebra, analysis, functional analysis, topology, probability, differential equations, etc. can be found (to within isomorphism) in $\mathcal{Z}$* . Many fundamental objects of abstract set theory are also in $\mathcal{Z}$, all we have constructed before this chapter in developing the theory of inductive posets, well ordered sets, etc. In slogan form: **we can develop classical mathematics and all the set theory needed for it as if all mathematical objects were members of $\mathcal{Z}$.**

The same can be said of every Zermelo universe, of course, but the concrete, simple definition of $\mathcal{Z}$ makes it possible to analyze its structure and investigate the special properties of its members. For example, no set which is a member of itself belongs to $\mathcal{Z}$: because no $X \in N_0$ satisfies $X \in X$ (easily), and if n were least such that some $X \in X \in \mathcal{Z}_{n+1}$, then $X \in X \subseteq \mathcal{Z}_n$ by the definition, so $X \in \mathcal{Z}_n$, contradicting the choice of n . This looks good, we had some trouble with sets which belong to themselves. Actually, the iterative construction of $\mathcal{Z}$ ensures a much stronger regularity property for its members, discovered by von Neumann.

11.26. Definition. An object x is **ill founded** if it is the beginning of a descending $\in$ -chain, i.e. if there exists a function $f : N \rightarrow E$ such that

$$x = f(0) \ni f(1) \ni f(2) \ni \cdots$$

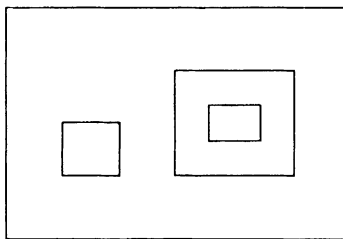


Figure 11.2. $\{\emptyset, \{\emptyset\}\}$ as a disappointing gift.

Objects which are not ill founded are **well founded** or **grounded**. If $X \in X$, then $X \ni X \ni X \ni \dots$, so X is ill founded. Problem **x11.14** gives a simple characterization of ill founded sets directly in terms of the $\in$ relation, which suggests that ill foundedness is a generalization of self-membership.

11.27. Exercise. *Atoms are grounded, as is $\emptyset$ and N_0 . A set is grounded if and only if all its members are grounded, if and only if its powerset is grounded. The class of all grounded sets is transitive.*

11.28. Proposition. *If I is grounded, then so is its basic closure $M(I)$. In particular, the least Zermelo universe $\mathcal{Z}$ and all its members are grounded.*

Proof. Assume that I is grounded, let (towards a contradiction) n be least such that M_n is ill founded and suppose that $M_n \ni x_1 \ni \dots$ is a descending $\in$ -chain. By hypothesis $n > 0$. Since $x_1 \in M_{n-1}$ and $x_1 \in y \in M_{n-1}$, contradict the choice of n , we must have $x_1 \subseteq M_{n-1}$, so $x_2 \in M_{n-1}$ and the descending $\in$ -chain $M_{n-1} \ni x_2 \ni \dots$ contradicts again the choice of n . It follows that M is also grounded, since any descending chain $M \ni x_1 \ni \dots$ would also witness that M_n is ill founded for whatever M_n contains x_1 . The consequence about $\mathcal{Z}$ follows because N_0 is grounded. $\dashv$

In the old gag, the excited birthday boy opens up the huge box with his present, only to find inside it another box, and inside that another, and so on, until the last, tiny box is empty: his present is just the boxes. We can think of a pure, grounded set as a disappointing gift of this sort, except that each box may contain several boxes, not just one; no matter which one the birthday boy chooses to open up each time, eventually he finds the empty box, $\emptyset$. Most axiomatizations of set theory ban ill founded sets from the start by adopting the following principle proposed by von Neumann.

11.29. Principle of Foundation. *Every set is grounded.* This is also called the principle (or axiom) of *Regularity* in the literature.

It is worth putting down here an equivalent version of this Principle,

which is somewhat opaque but useful.

11.30. Proposition. *The Principle of Foundation is true if and only if for every non-empty set X , there is some $m \in X$ such that*

$$m \cap X = \emptyset. \quad (11.24)$$

Proof. Assume first the Principle of Foundation and suppose, towards a contradiction, that $X \neq \emptyset$ but no $m \in X$ satisfies (11.24). This means that for some $a \in X$,

$$a \in X \ \& \ (\forall m \in X)(\exists t \in X)[t \in m],$$

and then **DC** gives us an infinite descending $\in$ -chain beginning with $X \ni a$ which contradicts the hypothesis. Conversely, if the Principle of Foundation fails and some infinite descending $\in$ -chain starts with some set

$$X = f(0) \ni f(1) \ni f(2) \ni \dots,$$

then the set $f[N] = \{f(0), f(1), \dots\}$ is not empty and intersects each of its members, so none of them satisfies (11.24). $\neg$

11.31. Axiomatics. By far the most widely used—the “official”—system of axioms for sets is the **Zermelo-Fraenkel Theory** (with choice), which accepts the Principles of Purity **3.24** and Foundation in addition to those of **ZFAC**, symbolically,

$$\mathbf{ZFC} = \mathbf{ZFDC} + \mathbf{AC} + \text{Purity} + \text{Foundation}.$$

There are many and convincing arguments in favor of this industry standard, some of which we discuss immediately below. We will come back to the question in Chapter 12 and Appendix B, where we will also explain a few good foundational reasons for sticking with the weaker **ZFDC** in these Notes. As a practical matter, the principles of Purity and Foundation do not come up in the part of the subject we are covering, and the full **AC** is only needed rarely, so we can easily keep track of it.

11.32. Are all sets grounded? The most blatant exception to the Principle of Foundation would be a set which is its own singleton,

$$\Omega = \{\Omega\}. \quad (11.25)$$

We can think of Ω as the ultimately frustrating gift: each box has exactly one box inside it, identical with the one you just opened, and you can keep opening them forever without ever finding anything. How about sets Ω^1 and Ω^2 such that

$$\Omega^1 = \{\emptyset, \Omega^2\}, \quad \Omega^2 = \{\Omega^1\}? \quad (11.26)$$

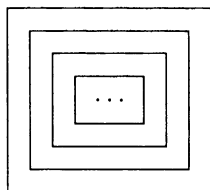


Figure 11.3. Ω as the ultimately frustrating gift.

These equations look unlikely, even bizarre, but it is not clear that our axioms rule them out. As a matter of fact they do not, we will construct in Appendix B some quite reasonable models of **ZFAC** which contain some $\Omega = \{\Omega\}$ and many other sets with similar properties.

Recall the discussion about the *large* and the *small* heuristic views of the universe of objects $\mathcal{W}$ in 3.25. The large view conceives $\mathcal{W}$ as the largest possible collection of objects which satisfies the axioms, while the small view takes it to comprise just the objects guaranteed by them.

On the large view, we have no more evidence in favor or against the Principle of Foundation now than we did back in Chapter 3, except that we have proved all these things about sets without ever using it. But then again, we never saw a need for ill founded sets either.

On the small view, we have amassed some considerable evidence, at least for **ZAC**, and it is all in favor of the Principle of Foundation: we now have a precise idea of what sets are “guaranteed” by the axioms of **ZAC**, they are the members of $\mathcal{Z}$ and they are all pure and grounded. It may be argued that we did not build $\mathcal{Z}$ out of whole cloth, we worked within a “given” universe $\mathcal{W}$ of objects, in fact, we needed to assume that $\mathcal{W}$ satisfies the Axiom of Replacement in addition to the axioms of **ZAC**. This is certainly true, but so is the obvious response to it: aside from any rigorous axiomatization, the definition of $\mathcal{Z}$ and the proofs of its basic properties can be understood intuitively, naively, and they carry a considerable force of persuasion. An informal description of $\mathcal{Z}$ would have made perfect sense in Chapter 3, as an intuitive conception of “restricted set” which justifies the axioms of **ZAC**, the principles of Purity and Foundation among them. We have not been able to produce any such plausible, intuitive model of **ZDC** which contains ill founded sets from any hypotheses which do not beg the question.³

Could we construct simple models like $\mathcal{Z}$ for the theory **ZFDC**? Let’s first give them a name.

³Models like $M(N_0 \cup \Omega)$ beg the question, because they need some Ω with the requisite self-membership property to get started.

11.33. Definition. A **Z-F universe** is any Zermelo universe M which further satisfies the Axiom of Replacement in the following sense: for each $A \in M$ and each unary definite operation H ,

$$(\forall x \in M)[H(x) \in M] \implies H[A] = \{H(x) \mid x \in A\} \in M.$$

The axioms of **ZFDC** assert precisely that the class $\mathcal{W}$ of all objects is a Z-F universe.

11.34. Theorem. *The von Neumann class*

$$\mathcal{V} =_{\text{df}} \{X \mid X \text{ is a pure, grounded set}\} \quad (11.27)$$

is a Z-F universe.

Proof. The fact that $\mathcal{V}$ is a Zermelo universe is quite trivial, most of it follows from Exercise 11.27. To verify that $\mathcal{V}$ also satisfies the Axiom of Replacement, notice that (whether $A \in \mathcal{V}$ or not), if H is unary, definite and such that for every $x \in A$, the value $H(x)$ is a pure, grounded set, then the image $H[A]$ has only pure and grounded members, so it is (easily) pure and grounded. $\dashv$

There is another, elegant and useful characterization of the pure grounded sets which follows easily from the Grounded Recursion Theorem 11.5.

11.35. Definition. A **Mostowski surjection or decoration** of a graph G with edge relation $\rightarrow$ is a surjection $d : G \rightarrow d[G]$ which assigns a set to each node of G such that

$$d(x) = \{d(y) \mid y \leftarrow x\} \quad (x \in G). \quad (11.28)$$

11.36. Mostowski Collapsing Lemma. (1) *Every grounded graph G admits a unique decoration d_G , and its image $d_G[G]$ is a transitive, pure, grounded set.*

(2) *A set A is pure and grounded if and only if there exists a grounded graph G and a node $x \in G$, such that $A = d_G(x)$ for the unique decoration d_G of G .*

Proof. (1) The existence of a unique decoration of a grounded G follows immediately from the Grounded Recursion Theorem 11.5 applied to G , with the definite operation

$$H(f) = \text{Image}(f) = \{f(x) \mid f(x) \downarrow\}.$$

The image $d_G[G]$ is transitive, since if $s \in t \in d_G[G]$, then $s \in t = d_G(y)$ for some $y \in G$, and then $s = d_G(x)$ for some $x \leftarrow y$, so $s \in d_G[G]$. Since each

$d_G(x)$ is a set, by (11.28), $d_G[G]$ is a transitive set with no atoms and hence pure. Finally, $d_G[G]$ is grounded, because if $x_0 \ni x_1 \ni \dots$ were an infinite, descending $\in$ -chain in it and $s_0, s_1, \dots$ were chosen so that $d_G(s_i) = x_i$, then $s_0 \rightarrow s_1 \rightarrow \dots$ would be an infinite descending chain in the grounded graph G .

(2) If $A = d_G(x)$ with x a node in some grounded graph, then A is a member of a transitive, pure, grounded set by (1) and hence pure and grounded. For the converse, let $G = TC(A)$ be the transitive closure of X and define on it

$$x \rightarrow y \iff_{\text{df}} y \in x.$$

The graph G is grounded, because $G = TC(A)$ is a grounded set, Problem **x11.16**. In addition,

$$d_G(x) = x \quad (x \in G); \quad (11.29)$$

because if x were a G -minimal counterexample to (11.29), then

$$\begin{aligned} d_G(x) &= \{d_G(y) \mid y \leftarrow_G x\}, \\ &= \{y \mid y \leftarrow x\} && \text{by the choice of } x, \\ &= \{y \mid y \in x\} && \text{by the def. of } \rightarrow, \\ &= x && \text{because } x \text{ is a set.} \end{aligned}$$

In particular, $A = d_G(A)$, which proves (2). $\dashv$

The class $\mathcal{V}$ is not a set (Problem **x11.20**) and it is quite hard to find Z-F universes which are sets. See Problems ***x12.20**, ***x12.21** and ***xB.12**.

11.37. Consistency and independence results. All the consistency and independence results we have discussed in **8.22**, **8.23**, **10.27**, **10.28**, **10.30** and **10.31** can be strengthened by adding the Axiom of Replacement to the relevant theories. This is as good a place as any to collect the most general versions of these fundamental results, which are outside the scope of these Notes.

(1) (Gödel, 1939) *The universe L of constructible sets is a model of **ZFC**, which further satisfies the Generalized Continuum Hypothesis **GCH**. It follows that the Axiom of Choice **AC** cannot be refuted from the other axioms of **ZFC**, and that **GCH** cannot be refuted in **ZFC**.*

(2) (Cohen, 1963) *None of the choice principles **AC**_N, **DC** and **AC** can be proved from a weaker one using the constructive axioms of Zermelo (I) - (VI) and the Axiom of Replacement (VIII).*

(3) (Cohen, 1963) *There is a model of **ZFC** in which the Continuum Hypothesis **CH** is false, so **CH** cannot be proved in **ZFC**.*

(4) (Solovay, 1970) *There is a model of **ZFC** in which every “definable”, uncountable pointset has a perfect subset, and hence has cardinality $\mathfrak{c}$. This*

means, in particular, that we cannot define a specific pointset A and then prove in **ZFC** that it has cardinality intermediate between $\aleph_0$ and $\mathfrak{c}$.

(5) (Solovay, 1970) *There is a model of ZFDC in which every uncountable pointset has a perfect subset, so we cannot prove in ZFDC the existence of uncountable pointsets without perfect subsets.* Solovay's model also satisfies the Principles of Purity and Foundation.

Problems

x11.1. Prove the Separation Axiom (**III**) from the remaining axioms in the group (**I**) - (**V**) and the Axiom of Replacement (**VIII**).

x11.2. For each set I , each unary definite operation F and each binary, definite operation G , there exists a least under $\subseteq$ set $\bar{A}$ which contains A as a subset and is closed under F and G , i.e.

$$A \subseteq \bar{A}, \quad x \in \bar{A} \implies F(x) \in \bar{A}, \quad x, y \in \bar{A} \implies G(x, y) \in \bar{A}.$$

(The same is true for any number of operations, of any number of arguments.)

x11.3. The Axiom of Replacement is constructively equivalent with the following proposition: for every set A and every unary definite operation F , there exists a set B which contains A and is closed under F , i.e.

$$A \subseteq B \ \& \ F[B] \subseteq B.$$

x11.4. The Axiom of Replacement is constructively equivalent with the following proposition: for every set A and every unary definite operation F , the restriction

$$F \upharpoonright A =_{\text{df}} \{(x, F(x)) \mid x \in A\}$$

of F to A is a function, i.e. a set of pairs.

x11.5. If (x, y) is a definite, binary operation which satisfies the first property of ordered pairs **4.1**, then it also satisfies the second, **4.2**. (This cannot be proved in **ZAC**, see Problem ***xB.4**.)

x11.6. If $|A|$ is a definite operation which satisfies the first condition on weak cardinal assignments (4.25), then it automatically also satisfies the third one, (4.27). (This cannot be proved in **ZAC**, see Problem ***xB.8**.)

x11.7. The definite condition of functionhood defined in (4.22) satisfies the equivalence

$$\begin{aligned} \text{Function}(f) \iff & \text{Set}(f) \ \& \ (\forall w \in f)(\exists x, y)[w = (x, y)] \\ & \& \ (\forall x, y, y')[[(x, y) \in f \ \& \ (x, y') \in f] \implies y = y'], \end{aligned}$$

i.e. f is a function exactly when it is a single-valued set of pairs. (See also Problem *xB.9.)

x11.8. There exists a unique sequence $(n \mapsto \aleph_n)$ which satisfies the identities

$$\aleph_0 = |\mathcal{N}|, \quad \aleph_{n+1} = \aleph_n^+.$$

We introduced these names for the first few infinite cardinals in (9.6), but this is not the same as proving the existence of *the sequence* $(n \mapsto \aleph_n)$. (See also Problem *xB.10.)

x11.9. Extended recursion with parameters. For every unary definite operation G and every binary definite operation H , there exists a unary definite operation F which satisfies the identities

$$\begin{aligned} F(0, y) &= G(y), \\ F(n+1, y) &= H(F(n, y), y). \end{aligned}$$

x11.10. If $\mathcal{E}$ is a non-empty family of transitive sets, then the union $\bigcup \mathcal{E}$ and the intersection $\bigcap \mathcal{E}$ are also transitive.

x11.11. For every class A there exists a least, transitive class $\overline{A}$ which contains A , that is, such that $A \subseteq \overline{A}$ and for every transitive class B , $A \subseteq B \implies \overline{A} \subseteq B$.

x11.12. The class of all pure sets is transitive, as are the classes of hereditarily finite and hereditarily countable sets.

x11.13. If $x_1 \in x_2 \in \cdots \in x_n = x_1$, then x_1 is ill founded.

x11.14. An object x is ill founded if and only if there exists some set A such that

$$x \in A \ \& \ (\forall s \in A)(\exists t \in A)[s \ni t].$$

x11.15. If Ω_1 and Ω_2 exist which satisfy (11.26), then they are distinct, hereditarily finite, pure sets.

x11.16. A set A is grounded if and only if its transitive closure $TC(A)$ is grounded.

***x11.17.** A set is in HF if and only if it is pure, grounded and hereditarily finite. HINT: Show first that every finite, transitive, pure grounded set is in HF . The members of HF are usually called *the hereditarily finite sets*, because the principles of Purity and Foundation are typically included among the axioms.

x11.18. For each transitive set I , let

$$J = \{x \in I \mid x \text{ is pure and grounded}\}$$

and prove that

$$M(J) = \{x \in M(I) \mid x \text{ is pure and grounded}\}.$$

x11.19. If a set $\Omega = \{\Omega\}$ exists as in (11.25), then

$$\{x \in M(\Omega) \mid x \text{ is grounded}\} = HF.$$

x11.20. Prove that the class $\mathcal{V}$ of all pure, grounded sets is not a set.

x11.21. A class K of atoms **supports** a set A if

$$x \in TC(A) \ \& \ Atom(x) \implies x \in K.$$

Prove that the class of sets supported by a fixed K is a Z-F universe.

x11.22. The class $\mathcal{V}[K]$ of grounded sets supported by a class K of atoms is a Z-F universe.

x11.23. An *extended decoration* of a graph G is any surjection $d : G \rightarrow d[G]$ such that for all $x \in G$,

$$d(x) = \begin{cases} x, & \text{if } x \text{ is an atom,} \\ \{d(y) \mid y \leftarrow x\}, & \text{if } x \text{ is a set.} \end{cases}$$

Prove that every grounded graph admits a unique extended decoration and that a (not necessarily pure) set A is grounded if and only if there exists a grounded graph G and some $x \in G$, such that $A = d(x)$.

***x11.24. Grounded $\in$ -recursion.** For each unary definite operation H , there exists a definite operation $F(t)$, such that for every grounded set x ,

$$F(x) = H(F \upharpoonright x),$$

where the function $F \upharpoonright x = \{(t, F(t)) \mid t \in x\}$ is the restriction of the operation F to the set x .

This is a special case of the next, slightly more complex generalization of the Grounded Recursion Theorem **11.5**.

***x11.25.** Suppose $t < x$ is a binary definite condition which satisfies (1) for every x , the class $\{t \mid t < x\}$ is a set, and (2) there does not exist a sequence $(n \mapsto x_n)$ such that for all n , $x_{n+1} < x_n$. Prove that for every definite operation H there exists another F , such that for every x ,

$$F(x) = H(F \upharpoonright \{t \mid t < x\}),$$

where the function $F \upharpoonright \{t \mid t < x\} = \{(t, F(t)) \mid t < x\}$ is the restriction of F to the set $\{t \mid t < x\}$.

Chapter 12

ORDINAL NUMBERS

The Axiom of Replacement finds its most important applications in von Neumann's beautiful theory of ORDINAL NUMBERS, and in the construction of the CUMULATIVE HIERARCHY of pure, grounded sets. One can live without knowing the ordinals, to be sure, but not as well: they bring many gifts, among them true cardinal numbers which give substance to the "virtual" theory of equinumerosities with which we have been making do. The Cumulative Hierarchy extends the iteration of the power operation we have used to construct HF "as far as it will go" and presents the pure, grounded sets as the most compelling intuitive understanding of what sets really are. It is not so clear one can live without knowing that, not among set theorists, at any rate.

Cantor describes his conception of "ordinal types" just a few pages after the definition of cardinals quoted in 4.19, and in a very similar vein.

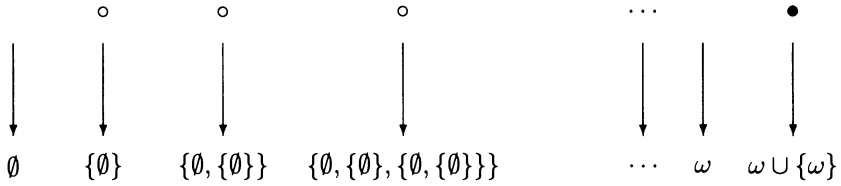
Every ordered set U has a definite 'ordinal type,' ... which we will denote by $\bar{U}$. By this we understand the general concept which results from U if we only abstract from the nature of the elements u , and retain the order or precedence among them. Thus the ordinal type $\bar{U}$ is itself an ordered set whose elements are units which have the same order of precedence amongst one another as the corresponding elements of U , from which they are derived by abstraction. ... A simple consideration shows that two ordered sets have the same ordinal type if, and only if, they are similar, so that of the two formulas $U =_o V$, $\bar{U} = \bar{V}$, one is always a consequence of the other.

Cantor is speaking about arbitrary linearly ordered sets, but we will consider here only the problem of defining "ordinal types" for well ordered sets. He states explicitly the first key property

$$U =_o \bar{U} \tag{12.1}$$

of the ordinal assignment operation and argues for

$$U =_o V \implies \bar{U} = \bar{V}. \tag{12.2}$$

**Figure 12.1.** The von Neumann surjection.

Cantor's implied "simple consideration" for (12.2) should also justify (for well ordered sets) the stronger implication

$$U \leq_o V \implies \bar{U} \subseteq \bar{V}; \quad (12.3)$$

because the position of a point x in a well ordered set depends only on the points preceding it, so the "unit" $\bar{x}$ abstracted from x and coding its place in U should depend only on the initial segment $\text{seg}_U(x)$. Thus, the problem of representing Cantor's conception of ordinals in axiomatic set theory comes down to the following: *can we assign a well ordered set $\bar{U}$ to each well ordered set U , so that (12.1) and (12.3) hold?* Von Neumann's ingenious idea is to define $\bar{U}$ by replacing *recursively* each member of U by the set of its predecessors.

12.1. Ordinal Numbers. *The von Neumann surjection of a well ordered set U is the unique surjection $\mathbf{v} = \mathbf{v}_U : U \twoheadrightarrow \mathbf{v}_U[U]$ which satisfies the identity*

$$\mathbf{v}(y) = \{\mathbf{v}(x) \mid x < y\} = \mathbf{v}[\text{seg}(y)] \quad (y \in U). \quad (12.4)$$

*We define the **ordinal number** of U to be the image*

$$\text{ord}(U) =_{\text{df}} \mathbf{v}_U[U] \quad (12.5)$$

of U under its von Neumann surjection, and we set

$$ON(\alpha) \iff \alpha \in ON \iff_{\text{df}} (\exists \text{ well ordered } U)[\alpha = \text{ord}(U)]. \quad (12.6)$$

12.2. Exercise. *Verify that (12.4) is equivalent to $\mathbf{v}(y) = H(\mathbf{v} \upharpoonright \text{seg}(y))$ where H is the definite operation $H(w) = \text{Image}(w)$, set to $\emptyset$, as usual, when w is not a function.*

Suppose for example that

$$U : 0_U, 1_U, 2_U, \dots, \omega_U, S_U(\omega_U)$$

is a well ordered set with least element 0_U , next $1_U, \dots$, first limit point ω_U , followed by the last (largest) point $S_U(\omega_U)$. We compute the values of

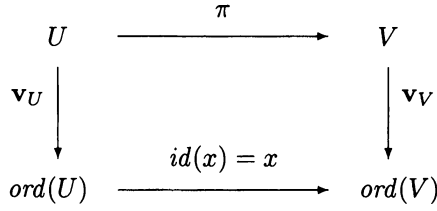


Figure 12.2. The von Neumann surjection under initial similarities.

its von Neumann surjection by repeated applications of (12.4):

$$\begin{aligned}
 \mathbf{v}(0_U) &= \{\mathbf{v}(x) \mid x < 0_U\} &= \emptyset &= 0, \\
 \mathbf{v}(1_U) &= \{\mathbf{v}(x) \mid x < 1_U\} &= \{\emptyset\} &= 1, \\
 \mathbf{v}(2_U) &= \{\mathbf{v}(x) \mid x < 2_U\} &= \{\emptyset, \{\emptyset\}\} &= 2, \\
 \mathbf{v}(3_U) &= \{\mathbf{v}(x) \mid x < 3_U\} &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} &= 3, \\
 &\vdots && \\
 \mathbf{v}(\omega_U) &= \{\mathbf{v}(x) \mid x < \omega_U\} &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \dots\} &= \omega \\
 \mathbf{v}(S_U(\omega_U)) &= \{\mathbf{v}(x) \mid x < S_U(\omega_U)\} &= \omega \cup \{\omega\} \\
 \mathbf{v}[U] &= \{0, 1, 2, \dots, \omega, \omega \cup \{\omega\}\}.
 \end{aligned}$$

Notice that each value $\mathbf{v}(y)$ is independent of the particular element $y \in U$, it depends only on the place of y in U , whether it is the first element, the fifth, the first limit point or whatever. This is a general fact about $\mathbf{v}$ which we can make precise as follows.

12.3. First Ordinal Property. *If $\pi : U \rightarrow \pi[U] \subseteq V$ is an initial similarity from a well ordered set U into another, then the diagram in Figure 12.2 commutes, i.e.*

$$\mathbf{v}_V(\pi(y)) = \mathbf{v}_U(y) \quad (y \in U). \quad (12.7)$$

Proof. By transfinite induction on U , we compute:

$$\begin{aligned}
 \mathbf{v}_V(\pi(y)) &= \{\mathbf{v}_V(t) \mid t <_V \pi(y)\} && \text{by definition,} \\
 &= \{\mathbf{v}_V(\pi(x)) \mid x <_U y\} && \text{because } \pi \text{ is initial,} \\
 &= \{\mathbf{v}_U(x) \mid x <_U y\} && \text{by the ind. hyp.} \\
 &= \mathbf{v}_U(y). && \dashv
 \end{aligned}$$

12.4. Exercise. *Prove that the object $\omega = \mathbf{v}_U(\omega_U)$ assigned to the first limit point of every well ordered set U (which has one) by its von Neumann surjection is the set*

$$\omega = \bigcap \{X \mid \emptyset \in X \ \& \ (\forall \alpha \in X)[\alpha \cup \{\alpha\} \in X]\}. \quad (12.8)$$

12.5. Second Ordinal Property. *For each well ordered set U and each $y \in U$,*

$$\mathbf{v}_U(y) = \text{ord}(\text{seg}_U(y)). \quad (12.9)$$

As a consequence, each von Neumann value $\mathbf{v}_U(y)$ is an ordinal number, and conversely, each ordinal number α is the von Neumann value $\mathbf{v}_U(y)$ of some point in a well ordered set. Put another way: each member of an ordinal is an ordinal and every ordinal is a member of an ordinal.

Proof. The key identity (12.9) says directly that von Neumann values are ordinals, and also the converse, since each $U = \text{seg}_{\text{Succ}(U)}(t)$, where $\text{Succ}(U)$ is the next well ordered set to U , with t added on top, 7.16. To prove (12.9), apply 12.3 with $\pi = (x \mapsto x)$ the identity on $W = \text{seg}_U(y)$, which is certainly an initial similarity of W into U . We get

$$\mathbf{v}_W(t) = \mathbf{v}_U(\pi(t)) = \mathbf{v}_U(t) \quad (t <_U y),$$

and by the definition,

$$\mathbf{v}_U(y) = \{\mathbf{v}_U(t) \mid t <_U y\} = \{\mathbf{v}_W(t) \mid t <_U y\} = \text{ord}(W). \quad \dashv$$

Thus, we can think of ordinal numbers as standing either for lengths of well ordered sets, or for places of points in a well ordered set. The latter agrees more with the use of ordinals in ordinary language, where “first,” “second,” ... customarily describe the place of objects in a sequence.

Next comes the basic fact about ordinal numbers.

12.6. Third Ordinal Property. *Each ordinal number α is well ordered by the relation*

$$u \leq_\alpha v \iff_{\text{df}} u = v \vee u \in v \quad (u, v \in \alpha), \quad (12.10)$$

and if $\alpha = \text{ord}(U)$ for a well ordered set U , then the von Neumann surjection $\mathbf{v} : U \rightarrow \alpha$ is a similarity.

Proof. The crucial implication is

$$\mathbf{v}(x) \in \mathbf{v}(y) \implies x < y \quad (x, y \in U) \quad (12.11)$$

where $\mathbf{v} : U \rightarrow \alpha = \text{ord}(U)$ is the von Neumann surjection of any well ordered set U . Together with its converse

$$x < y \implies \mathbf{v}(x) \in \mathbf{v}(y) \quad (x, y \in U)$$

which follows from the definition of $\mathbf{v}$, it implies directly that $\mathbf{v} : U \rightarrow \alpha$ is a bijection satisfying

$$x \leq y \iff \mathbf{v}(x) \leq_\alpha \mathbf{v}(y) \quad (x, y \in U),$$

which in turn implies immediately that $\leq_\alpha$ well orders α and $\mathbf{v}$ is a similarity of U with the well ordered set $(\alpha, \leq_\alpha)$.

To prove (12.11) by contradiction, suppose x is least in U such that for some y ,

$$\mathbf{v}(x) \in \mathbf{v}(y), \quad y \leq x. \quad (12.12)$$

We take cases on the inequality $y \leq x$.

CASE 1. $y = x$, so $\mathbf{v}(x) \in \mathbf{v}(x)$, hence $\mathbf{v}(x) = \mathbf{v}(s)$ for some $s < x$ and $\mathbf{v}(s) \in \mathbf{v}(s)$, which contradicts the choice of x . (Because with $t = s$, we have $\mathbf{v}(s) \in \mathbf{v}(t)$ and $t \leq s$.)

CASE 2. $y < x$. We now have both $\mathbf{v}(x) \in \mathbf{v}(y)$ and $\mathbf{v}(y) \in \mathbf{v}(x)$, so by the definitions we get $s < x$ and $t < y$ such that $\mathbf{v}(y) = \mathbf{v}(s)$ and $\mathbf{v}(x) = \mathbf{v}(t)$, so

$$\mathbf{v}(t) \in \mathbf{v}(s), \quad \mathbf{v}(s) \in \mathbf{v}(t). \quad (12.13)$$

Both s and t are below x , since for the latter, $t < y \leq x$, and (12.13) implies that whichever of s, t is smaller than the other contradicts the choice of x . $\neg$

12.7. Corollary. (1) *Every well ordered set is similar with an ordinal number.*

(2) *Every well orderable set is equinumerous with an ordinal number.*

Proof. Part (1) restates 12.6 and Part (2) follows because similarities are bijections. $\neg$

This remarkable result says, in effect, that there exist sufficiently long $\in$ -chains to mirror every wellordering and it is a characteristic consequence of the Replacement Axiom. As we have been doing with structured sets throughout, by “the ordinal α ” we will mean ambiguously the set α , e.g. in (2) of 12.7, or the well ordered set $(\alpha, \leq_\alpha)$, e.g. in (1) of 12.6. Notice that in this case the set α determines the ordering $\leq_\alpha$ by (12.10).

The three properties of ordinals 12.3, 12.5 and 12.6 give a strong solution to Cantor’s problem of defining ordinal types for well ordered sets, as we described it above.

12.8. Theorem. *The definite operation $\text{ord}(U)$ on well ordered sets satisfies the conditions*

$$U =_o \text{ord}(U), \quad (12.14)$$

$$U \leq_o V \implies \text{ord}(U) \subseteq \text{ord}(V), \quad (12.15)$$

$$ON(\alpha) \implies \alpha = \{\beta \in ON \mid \beta <_o \alpha\}. \quad (12.16)$$

Proof. The first property (12.14) is a restatement of part of 12.6. To prove (12.15), suppose $\pi : U \rightarrow \pi[U] \subseteq V$ is an initial similarity. By 12.3, taking images, $\mathbf{v}_V[\pi[U]] = \mathbf{v}_U[U]$; since $\mathbf{v}_V$ is a similarity of V with $\text{ord}(V)$, it carries initial segments onto initial segments, so $\mathbf{v}_U[U] = \mathbf{v}_V[\pi[U]] \subseteq \text{ord}(V)$. Finally, for (12.16), if $\alpha = \text{ord}(U)$, then:

$$\begin{aligned} \alpha &= \{\mathbf{v}_U(y) \mid y \in U\} && \text{by def.} \\ &= \{\text{ord}(\text{seg}_U(y)) \mid y \in U\} && \text{by (12.5)} \\ &= \{\beta \in ON \mid \beta <_o \alpha\}, \end{aligned}$$

the last because the well ordered sets which are $<_o U$ are exactly those similar with the proper initial segments of U . $\dashv$

Conditions (12.14) and (12.15) are precisely Cantor's (12.1) and (12.3). The key, last condition (12.16) is characteristic of the von Neumann ordinal assignment and ensures its uniqueness, Problem *x12.1. This is an interesting result, we formulated it as a problem only because it makes for a good one and we will not need to appeal to it. However, it is easy to get lost in proving scores of elementary properties of ordinals, some useful, others just challenging, and the proofs from the definition are a bit confusing: it is not entirely natural to think of the membership relation as an ordering. It is good practice, at least in the beginning, to prove properties of ordinals directly from conditions (12.14) - (12.16).

12.9. Theorem. Characterization of ordinals. *A set α is an ordinal if and only if it is transitive, pure, and well ordered by the relation*

$$x \leq_\alpha y \iff_{\text{df}} x = y \vee x \in y \quad (x, y \in \alpha), \quad (12.17)$$

equivalently, if α is transitive, grounded, pure and connected, i.e.

$$x, y \in \alpha \implies x \in y \vee x = y \vee y \in x. \quad (12.18)$$

Proof. For one direction, suppose $\alpha = \mathbf{v}[U]$ for some well ordered set U . Certainly α has no atoms, since every von Neumann value $\mathbf{v}(y)$ is a set, by its definition. If $t \in x \in \alpha$, then $t \in \mathbf{v}(v)$ for some $v \in U$ such that $\mathbf{v}(v) = x$, so $t = \mathbf{v}(u)$ for some $u < v$ by the definition of $\mathbf{v}(v)$ and hence $t \in \alpha$. This shows that α is transitive and it also satisfies the last condition by 12.6.

For the converse, suppose that α is a transitive set with no atoms, well ordered by $\leq_\alpha$ in (12.17). We prove by transfinite induction on α that for each $y \in \alpha$, $\text{ord}(\text{seg}_\alpha(y)) = y$:

$$\begin{aligned} \text{ord}(\text{seg}_\alpha(y)) &= \{\text{ord}(\text{seg}_\alpha(x)) \mid x <_\alpha y\} && \text{by (12.16),} \\ &= \{\text{ord}(\text{seg}_\alpha(x)) \mid x \in \alpha \ \& \ x \in y\} && \text{by def. of } <_\alpha \\ &= \{x \mid x \in \alpha \ \& \ x \in y\} = \alpha \cap y && \text{by ind. hyp.} \\ &= y, \end{aligned}$$

where $\alpha \cap y = y$ holds because $y \in \alpha$ and α is transitive with no atoms. (Where does the absence of atoms come in?) From this and 12.5 we get

$$\alpha = \{y \mid y \in \alpha\} = \{\text{ord}(\text{seg}_\alpha(y)) \mid y \in \alpha\} = \{v_\alpha(y) \mid y \in \alpha\} = \text{ord}(\alpha),$$

so that α is an ordinal.

The second characterization follows trivially. $\dashv$

Thus, in **ZFC**, ordinals are precisely the transitive sets which satisfy (12.18), which is the way they are often defined. It is a remarkably simple definition, but if you start with it, it takes some time and effort to see what it has to do with the faithful representation of well ordered sets and cardinals.

The simple, uniform definition of the orderings $\leq_\alpha$ in terms of $\in$ yields an equally simple characterization of the comparison of ordinals by initial similarities.

12.10. Theorem. *For every two ordinals α, β ,*

$$\alpha \leq_o \beta \iff \alpha = \beta \vee \alpha \in \beta \iff \alpha \sqsubseteq \beta \iff \alpha \subseteq \beta.$$

Proof. We give a round-robin argument of the strict versions of the claimed equivalences.

(1) $\alpha <_o \beta \implies \alpha \in \beta$ follows immediately from (12.14) and (12.16), since the hypothesis means that $\alpha = \text{ord}(U)$ and $\beta = \text{ord}(V)$ with $U <_o V$.

(2) $\alpha \in \beta \implies \alpha \sqsubset \beta$ and (3) $\alpha \sqsubset \beta \implies \alpha \subsetneq \beta$ follow equally easily from (12.14) and (12.16) and we will skip them.

It remains to show $\alpha \subsetneq \beta \implies \alpha <_o \beta$ to close the loop. Assume the hypothesis and let ξ be the $<_o$ -least member of $\beta \setminus \alpha$. Thus, for $\eta \in \beta$,

$$\eta \in \xi \implies \eta <_o \xi \implies \eta \in \alpha,$$

by (12.16) and the choice of ξ . But also

$$\eta \in \alpha \implies \eta <_o \xi \implies \eta \in \xi$$

similarly, so $\alpha = \xi \in \beta$, and by (12.16) again, $\alpha <_o \beta$. $\dashv$

It is traditional to use for the ordering on ordinals the simplest notation,

$$\alpha \leq \beta \iff_{\text{df}} \alpha \leq_o \beta \quad (\alpha, \beta \in ON), \quad (12.19)$$

keeping in mind its equivalent characterizations in (12.10). We summarize its properties in one, now simple result.

12.11. The order of ordinals. (1) *The class ON of ordinal numbers is well ordered by the condition $\alpha \leq \beta$, in the following precise sense:*

$$\begin{aligned} \alpha &\leq \alpha, \\ \alpha \leq \beta \ \&\ \beta \leq \gamma &\implies \alpha \leq \gamma, \\ \alpha \leq \beta \ \&\ \beta \leq \alpha &\implies \alpha = \beta, \\ \alpha < \beta \vee \alpha = \beta &\vee \beta < \alpha, \end{aligned}$$

and for every definite condition P ,

$$(\exists \alpha \in ON)P(\alpha) \implies (\exists \alpha \in ON)[P(\alpha) \ \&\ (\forall \beta < \alpha) \neg P(\beta)].$$

In particular, there is no infinite descending chain of ordinals,

$$\alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \cdots \implies (\exists n)[\alpha_n = \alpha_{n+1}]. \quad (12.20)$$

When $P(\alpha)$ holds for some α , we set

$$(\mu \alpha \in ON)P(\alpha) = \inf \{ \alpha \in ON \mid P(\alpha) \}. \quad (12.21)$$

(2) *For each ordinal number there is a next one,*

$$S(\alpha) =_{\text{df}} (\mu \beta \in ON)[\alpha < \beta] = \alpha \cup \{ \alpha \}. \quad (12.22)$$

(3) *Each set A of ordinal numbers has a least upper bound,*

$$\sup A =_{\text{df}} (\mu \beta \in ON)(\forall \alpha \in A)[\alpha \leq \beta] = \bigcup A, \quad (12.23)$$

which is the maximum of A if A has one and 0 if $A = \emptyset$.

Proof is left for Problem x12.2. \(\dashv\)

12.12. Exercise. *For each non-empty set of ordinals $\mathcal{E}$,*

$$(\mu \alpha \in ON)[\alpha \in \mathcal{E}] = \bigcap \mathcal{E}.$$

This follows from 12.10.

The **successor ordinals** are those of the form $S(\alpha)$ and the **limit ordinals** are those which are not successors or 0; they are obviously characterized by the property

$$\text{Limit}(\lambda) \iff \lambda \neq 0 \ \&\ \lambda = \sup \{ \alpha \mid \alpha < \lambda \}. \quad (12.24)$$

We can prove properties of ordinals by *transfinite induction* and define operations on them by *transfinite recursion*, as follows.

12.13. Induction on ON . For every unary definite condition P ,

$$(\forall \alpha)[(\forall \xi < \alpha)P(\xi) \implies P(\alpha)] \implies (\forall \alpha)P(\alpha).$$

Proof. Assume the hypothesis and that $\neg P(\alpha)$ for some α , let $\beta = (\mu \alpha \in ON) \neg P(\alpha)$ and derive a contradiction from $\neg P(\beta)$ and the choice of β . $\neg$

12.14. Recursion on ON . For every unary definite operation H , there exists another F , which satisfies the identity

$$F(\alpha) = H(F \upharpoonright \alpha) \quad (ON(\alpha)). \quad (12.25)$$

Here $F \upharpoonright \alpha$ is the function $\{(\xi, F(\xi)) \mid \xi \in \alpha\}$ obtained by restricting $F(\xi)$ to $\alpha = \{\xi \mid \xi < \alpha\}$.

Proof. For each β , by **11.6** on the well ordered set $(\beta, \leq_\beta)$, there exists exactly one function $f_\beta : \beta \rightarrow E_\beta$ which satisfies the identity

$$\begin{aligned} f_\beta(\alpha) &= H(f_\beta \upharpoonright \{x \in \beta \mid x <_\beta \alpha\}) \quad (\alpha < \beta), \\ &= H(f_\beta \upharpoonright \alpha), \end{aligned} \quad (12.26)$$

using the fact that $<_\beta$ coincides with $\in$ and the members of β are ordinals. We claim that

$$\alpha < \beta, \alpha < \gamma \implies f_\beta(\alpha) = f_\gamma(\alpha);$$

if not, there would exist a least α for which this fails for some β and γ , and then (12.26) yields a contradiction immediately. Thus, we can set

$$F(\alpha) = f_{S(\alpha)}(\alpha),$$

so $F(\alpha) = f_\beta(\alpha)$ for any $\beta > \alpha$ and (12.26) implies the required identity for $F(\alpha)$. $\neg$

Using this theorem, we can define arithmetical operations on ON and study their structure. We will leave most of this for the problems, but it is worth recording here the two most basic definitions, as examples of **12.14** and in order to have some notation available to name specific ordinals.

12.15. Ordinal addition and multiplication. There exist binary, definite operations $\alpha + \beta$ and $\alpha \cdot \beta$ on the ordinals which satisfy the following identities:

$$\begin{aligned} \alpha + 0 &= \alpha, \\ \alpha + S(\beta) &= S(\alpha + \beta), \\ \alpha + \lambda &= \sup \{\alpha + \beta \mid \beta < \lambda\}, \text{ if } \text{Limit}(\lambda). \end{aligned} \quad (12.27)$$

$$\begin{aligned} \alpha \cdot 0 &= 0, \\ \alpha \cdot S(\beta) &= (\alpha \cdot \beta) + \alpha, \\ \alpha \cdot \lambda &= \sup \{\alpha \cdot \beta \mid \beta < \lambda\}, \text{ if } \text{Limit}(\lambda). \end{aligned} \quad (12.28)$$

Proof. First we define by **12.14** a unary operation $+^\alpha$ for each α which satisfies the identities above, and then we set

$$\alpha + \beta =_{\text{df}} +^\alpha(\beta),$$

and similarly for multiplication. ⊢

We have already mentioned

$$\omega = \text{ord}(N), \quad (12.29)$$

the ordinal of the number sequence, characterized in (12.8). The ordinals following it immediately are obviously

$$\omega + 1 = S(\omega), \quad \omega + 2 = S(\omega + 1), \quad \omega + 3 = S(\omega + 2), \dots$$

and right above these comes

$$\omega + \omega = \sup \{ \omega + n \mid n \in \omega \} = \omega \cdot 2. \quad (12.30)$$

This is the second limit ordinal, the first one above ω . Each $\omega \cdot n$ can be obtained by adding ω to itself n times, directly from the definition. Next comes

$$\omega^2 = \sup \{ \omega \cdot n \mid n < \omega \},$$

after a while $\omega^3 = \omega^2 \cdot \omega$, etc.

Next we describe von Neumann's elegant solution of the problem of *cardinal assignment* **4.20** for well orderable sets, which is based on the fact that every one of them is equinumerous with an ordinal, **12.7**.

12.16. A von Neumann cardinal assignment is a definite operation $|A|$ on the class of sets which satisfies the following conditions:

$$A =_c |A|, \quad (12.31)$$

$$A =_c B \implies |A| =_c |B|, \quad (12.32)$$

$$(\forall \mathcal{E})[\{ |A| \mid A \in \mathcal{E} \} \text{ is a set}], \quad (12.33)$$

$$\text{if } A \text{ is well orderable, then } |A| = (\mu \kappa \in ON)[A =_c \kappa]. \quad (12.34)$$

The first three of these conditions characterize “weak” cardinal assignments in the terminology of **4.21**. The last condition (12.34) implies trivially that if A is well orderable, then for all B ,

$$A =_c B \iff |A| = |B|, \quad (12.35)$$

so that $|A|$ is a “strong” cardinal assignment by **4.21** on the class of well orderable sets. Of course, if **AC** holds, then every set is well orderable, so a Neumann cardinal assignment satisfies all of Cantor's conditions for an assignment of cardinal numbers to sets.

12.17. Theorem. *There exists a von Neumann cardinal assignment.*

Proof. We set

$$|A| = \begin{cases} (\mu\kappa \in ON)[A =_c \kappa], & \text{if } A \text{ is well orderable,} \\ A, & \text{otherwise.} \end{cases} \quad (12.36)$$

It is quite easy to verify the first two conditions (12.31) and (12.32), and the fourth one (12.34) is trivial, using (12.7). The third condition (12.33) follows from the first two, easily, by the Axiom of Replacement. $\dashv$

12.18. Cardinal numbers (2). We now assume that the cardinal assignment with which we have been working since 4.21 in Chapter 3 is, in fact, a von Neumann assignment, so that it satisfies (12.34) in addition to (12.31) - (12.33). The values of $|A|$ for well orderable A , the *well orderable cardinals* are also called **von Neumann cardinals**. Using (12.34) (easily), they are characterized as the *initial ordinals*, i.e.

$$Card_v(\kappa) \iff_{df} Card(\kappa) \ \& \ \kappa \text{ is well orderable,} \quad (12.37)$$

$$\iff \kappa \in ON \ \& \ (\forall \lambda < \kappa)[\lambda <_c \kappa], \quad (12.38)$$

Problem x12.11. By 9.11,

$$Card_v(\kappa) \implies Card_v(\kappa^+),$$

and by 9.18 and 9.20, if $\mathcal{E}$ is a non-empty set of cardinals, then

$$(\forall \kappa \in \mathcal{E}) Card_v(\kappa) \implies Card_v(\inf_c(\mathcal{E})) \text{ and } Card_v(\sup_c(\mathcal{E}));$$

in fact immediately from (12.34), (12.23) and 12.12, if $\mathcal{E}$ is a non-empty set of von Neumann cardinals, then

$$\inf_c(\mathcal{E}) = \bigcap \mathcal{E}, \quad \sup_c(\mathcal{E}) = \sup \mathcal{E} = \bigcup \mathcal{E}.$$

12.19. Cardinals, Choice and Replacement. One can make a good case that Cantor's *units* in the intuitive description of cardinals quoted in 4.19 are modeled faithfully by the von Neumann's ordinals, and the quotation

$\overline{\overline{A}}$ grows, so to speak, out of A in such a way that from every element x of A a special unit of A arises

describes precisely the construction of $|A| = ord(A) = \mathbf{v}[A]$ relative to some best wellordering of A , Problem x12.9. Whatever the value of the imagery, von Neumann's construction certainly makes it possible to delete the subscript $_c$ from all the equinumerosities we have established, applied

to von Neumann cardinals, so they become true identities of cardinal arithmetic. This is particularly useful if we work in **ZFAC**, which proves that all cardinals are von Neumann cardinals.

We have been careful to state results about cardinality without assuming the full Axiom of Choice whenever this was possible, but of course the main effect has been to make clear just how poor cardinal arithmetic is without it. The main problem is the equivalence of cardinal comparability with **AC**: we do not have much of an arithmetic if we cannot compare cardinals, and we cannot assume comparability without (necessarily) conceding the truth of the full Axiom of Choice.

Granting **AC**, how important is the existence of “true cardinals” which satisfy (12.35) and whose construction requires not only **AC** but also the Axiom of Replacement? Not much, by any account, unless you are allergic to subscripts. Thus, it might appear that von Neumann’s solution of the problem of Cardinal Assignment is primarily an exercise in mathematical elegance. There is some truth to this, but one must not draw the further conclusion that the Axiom of Replacement is unimportant for cardinal arithmetic, just because its basic identities can be established in **ZAC**, as equinumerosities. The problem is that **ZAC** cannot prove the existence of any cardinals above the first infinite sequence

$$\aleph_0, \aleph_1, \aleph_2, \dots,$$

and in fact it cannot even show that the sequence $(n \mapsto \aleph_n)$ exists, Problem ***xB.10**. In particular, the existence of singular cardinals cannot be shown in **ZAC**, so that the whole theory of cofinality remains possibly vacuous without Replacement.

The upshot is that to have a decent cardinal arithmetic, you must assume both the Axioms of full Choice and Replacement, i.e. to work in a theory at least as strong as **ZFAC**. It is sometimes assumed that the Principle of Foundation is also necessary for cardinal arithmetic, but this is not true—although some of the most important applications of cardinals are to the structure of von Neumann’s universe $\mathcal{V}$ of pure, grounded sets.

12.20. Proposition. *By recursion on $\alpha \in ON$, we set*

$$\begin{aligned} \aleph_0 &= |N| = \omega, \\ \aleph_{\beta+1} &= \aleph_\beta^+, \\ \aleph_\lambda &= \sup \{ \aleph_\beta \mid \beta < \lambda \}, \text{ if } \text{Limit}(\lambda). \end{aligned} \tag{12.39}$$

Each $\aleph_\alpha$ is a von Neumann cardinal,

$$\alpha < \beta \implies \aleph_\alpha <_c \aleph_\beta \quad (\alpha, \beta \in ON),$$

and every von Neumann cardinal is $\aleph_\alpha$ for some α .

Proof is simple enough to leave for a problem, **x12.12**. +

The operation $\aleph_\alpha$ supplies a useful notation for cardinal arithmetic, especially when we accept the Axiom of Choice.

12.21. Exercise. *The Axiom of Choice, AC, is equivalent to the proposition that “every cardinal is an aleph,”*

$$(AC) \quad (\forall A)(\exists \alpha)[A =_c |A| = \aleph_\alpha].$$

12.22. Exercise. (AC) *The Generalized Continuum Hypothesis is equivalent to the cardinal identity*

$$(GCH) \quad 2^{\aleph_\alpha} = \aleph_{\alpha+1} \quad (\alpha \in ON).$$

12.23. The Cumulative Hierarchy of Pure, Grounded Sets. *For each ordinal α we define the set $\mathcal{V}_\alpha$ by the following recursion on ON .*

$$\begin{aligned} \mathcal{V}_0 &= \emptyset, \\ \mathcal{V}_{\alpha+1} &= \mathcal{P}(\mathcal{V}_\alpha), \\ \mathcal{V}_\lambda &= \bigcup_{\alpha < \lambda} \mathcal{V}_\alpha, \quad \text{if } \text{Limit}(\lambda). \end{aligned}$$

The von Neumann universe is the union of all the $\mathcal{V}_\alpha$'s

$$\mathcal{V} =_{\text{df}} \bigcup_{\alpha \in ON} \mathcal{V}_\alpha = \{x \mid \text{for some } \alpha \in ON, x \in \mathcal{V}_\alpha\}, \quad (12.40)$$

and on it we define the rank operation by

$$\text{Rank}(x) = (\mu \alpha \in ON)[x \in \mathcal{V}_{\alpha+1}] \quad (x \in \mathcal{V}). \quad (12.41)$$

We have already used the symbol $\mathcal{V}$ to denote the class of pure grounded sets, because of the next result.

12.24. Theorem. (1) *Each $\mathcal{V}_\alpha$ is a pure, transitive, grounded set, and*

$$\alpha \leq \beta \implies \mathcal{V}_\alpha \subseteq \mathcal{V}_\beta.$$

(2) *If λ is a limit ordinal, $\lambda \geq \omega \cdot 2$, then $\mathcal{V}_\lambda$ is a Zermelo universe.*

(3) *For each pure set A , $A \subseteq \mathcal{V} \implies A \in \mathcal{V}$.*

(4) *The von Neumann universe $\mathcal{V}$ comprises the pure, grounded sets and is a Z-F universe.*

Proof. The arguments for Parts (1) and (2) are those we used to prove the corresponding properties of the basic closure sets $M(I)$ with transitive I in **x9.6**, and we will not repeat them.

(3) Assume that $A \subseteq \mathcal{V}$ and let $\text{Rank}[A] = \{\text{Rank}(x) \mid x \in A\}$ be the image of the rank operation on A . This is a set of ordinals, so there exists some ordinal κ strictly above its members,

$$x \in A \implies \text{Rank}(x) < \kappa \implies x \in \mathcal{V}_\kappa,$$

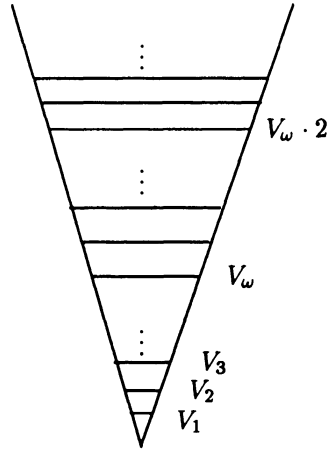


Figure 12.3. Logarithmic rendition of the pure, grounded sets.

and (using the purity of A) $A = \{x \in \mathcal{V}_\kappa \mid x \in A\} \in \mathcal{V}_{\kappa+1}$.

(4) We use the contrapositive of (3), which says that for pure A ,

$$A \notin \mathcal{V} \implies (\exists x \in A)[x \notin \mathcal{V}]. \quad (12.42)$$

Suppose first that M is pure, transitive and grounded but $M \notin \mathcal{V}$. The transitivity of M and (12.42) yield

$$(\forall x \in M, x \notin \mathcal{V})(\exists y \in M, y \notin \mathcal{V})[y \in x],$$

and then **DC** gives us a descending $\in$ -chain which proves M ill founded, contradicting the hypothesis. Thus, every pure, transitive, grounded set is in $\mathcal{V}$, so for every pure, grounded set A , $TC(A) \in \mathcal{V}$, and then $A \in \mathcal{V}$, since $A \in TC(A)$ and $\mathcal{V}$ is transitive. $\dashv$

12.25. The naive notion of pure, grounded set. In discussing the Principle of Foundation in 11.32, we argued that the definition of the least Zermelo universe $\mathcal{Z}$ and the proofs of its basic properties can be understood directly and naively, as we usually understand mathematics, and that they carry considerable force of persuasion as an intuitive conception of “set” which justifies the axioms of **ZAC** and the Principle of Foundation. In the same vein, we can argue that the “construction” of von Neumann’s universe $\mathcal{V}$ in 12.23 and 12.24 can be understood directly and naively outside the details of any specific axiomatization, and that it puts forward a natural, intuitive conception of “pure, grounded set” which justifies the axioms of **ZFC**. It is worth looking into these arguments a little closer.

The construction of $\mathcal{Z}$ begins with the infinite set

$$N_0 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots\}$$

and iterates the powerset operation $\mathcal{P}(A)$ infinitely many times. Since the “infinity” of iterations involved is no more and no less than that embodied by N_0 , we can say that *to understand $\mathcal{Z}$ we must understand two infinitary things: the set N_0 (basically the natural numbers) and the powerset operation.*

The construction of $\mathcal{V}$ starts with nothing, the empty set, but it proceeds to iterate the powerset operation $\mathcal{P}(A)$ through all the ordinals. On the same analysis, it is fair to say that *to understand $\mathcal{V}$ we must understand the class of ordinals ON and the powerset operation.* One may attempt to speak eloquently about the ordinals and justify them, as one might try to justify the natural numbers or the powerset operation. It should be clear, however, that the ordinals represent a separate and different new ingredient in our intuitive understanding of $\mathcal{V}$, they cannot be reduced to N_0 and the taking of powersets. From this point of view, the justification of the axioms of **ZFC** which we find in this intuitive construction is considerably weaker than the justification of **ZAC** we get from contemplating $\mathcal{Z}$.

In Appendix B we will consider alternative set universes, including some which contain both atoms and ill founded sets, and in more advanced textbooks one can find a multitude of fascinating models of set theory constructed (primarily) by extensions and combinations of Gödel’s constructibility and Cohen’s forcing. Part of the reason we have worked here in the weak systems of **ZDC** and **ZFDC** is to ensure that the elementary results of the field which we have covered apply directly to (essentially) all these models. These models, however, are all constructed starting with some given model of **ZFDC**, and it does not seem possible to produce for any of them independent, intuitive notions of **what sets are**, which justify directly the axioms they satisfy. It appears that (as of now), the intuitive conception of **pure, grounded set**, which is gleaned from an informal analysis of **12.23** and **12.24**, is by far the best replacement we have for Cantor’s unfettered (and self-contradictory) notion of “collection into a whole of definite and separate objects.”

Problems

***x12.1. Characterization of von Neumann ordinals.** Suppose $\phi(V)$ is a definite operation which assigns well ordered sets to well ordered sets and which satisfies the following three conditions:

$$\begin{aligned} V &=_{\circ} \phi(V), \\ U \leq_{\circ} V &\implies \phi(U) \subseteq \phi(V), \\ \text{Field}(\phi(V)) &= \{\text{Field}(\phi(U)) \mid U <_{\circ} V\}. \end{aligned}$$

Prove that $\text{ord}(V) = \alpha \implies \phi(V) = (\alpha, \leq_{\alpha})$.

x12.2. Prove **12.11**.

x12.3. The class ON is not a set.

x12.4. For every ordinal α ,

$$\alpha + 1 = \text{ord}(\text{Succ}(\alpha)),$$

by the definition of the successor poset $\text{Succ}(P)$ in **7.16**.

x12.5. For all ordinals α, β ,

$$\alpha + \beta = \text{ord}(\alpha +_o \beta),$$

by the definition of addition of posets in **7.37**. Infer from this that addition of ordinals is associative but not commutative, or give independent proofs of these facts.

x12.6. For all ordinals α, β ,

$$\alpha \cdot \beta = \text{ord}(\alpha \cdot_o \beta),$$

by the definition of multiplication of posets in **7.38**. Infer from this that multiplication of ordinals is associative but not commutative, or give independent proofs of these facts.

x12.7. Define an operation of exponentiation on the ordinals which (for $\alpha \neq 0$) satisfies the conditions

$$\begin{aligned} \alpha^0 &= 1, \\ \alpha^{S(\beta)} &= \alpha^\beta \cdot \alpha, \\ \alpha^\lambda &= \sup \{ \alpha^\beta \mid \beta < \lambda \}, \text{ if } \text{Limit}(\lambda). \end{aligned}$$

Which of the usual laws of exponents are valid for ordinal exponentiation? For example, is it always true

$$\alpha^{(\beta+\gamma)} = \alpha^\beta \cdot \alpha^\gamma?$$

***x12.8.** The only ordinals which belong to the least Zermelo universe $\mathcal{Z}$ are the finite ones.

x12.9. If $\leq$ is a best wellordering of A , then $|A| = \mathbf{v}_U[A]$, i.e. $|A|$ is the ordinal assigned to the well ordered set $(A, \leq)$ by its von Neumann surjection.

x12.10. The class Card_v of von Neumann cardinal numbers is not a set.

x12.11. Prove the characterization (12.37).

x12.12. Prove Proposition 12.20.

x12.13. (AC) The definite operation $\beth_\alpha$ is defined by the following recursion on ON :

$$\begin{aligned}\beth_0 &= \aleph_0 = |N| = \omega, \\ \beth_{\beta+1} &= 2^{\beth_\beta}, \\ \beth_\lambda &= \sup \{ \beth_\beta \mid \beta < \lambda \}, \text{ if } \text{Limit}(\lambda).\end{aligned}\tag{12.43}$$

Prove that for every ordinal α ,

$$|\mathcal{V}_{\omega+\alpha}| = \beth_\alpha.$$

12.26. Definition. A unary, definite operation F on the ordinals is **normal** if it is monotone,

$$\alpha \leq \beta \implies F(\alpha) \leq F(\beta),$$

and continuous at limit ordinals, i.e.

$$F(\lambda) = \sup \{ F(\beta) \mid \beta < \lambda \}, \text{ if } \text{Limit}(\lambda).$$

***x12.14.** Every normal operation on the ordinals has a fixed point, i.e. $F(\alpha) = \alpha$ holds for some α .

***x12.15. (AC)** There exist von Neumann cardinals κ, λ , such that

$$\kappa = \aleph_\kappa, \quad \lambda = \beth_\lambda.$$

12.27. Definition. Suppose $\lambda \leq \kappa$ are infinite limit ordinals. A function $f : \lambda \rightarrow \kappa$ is **cofinal** if it is strictly monotone, i.e.

$$\alpha < \beta < \lambda \implies f(\alpha) < f(\beta) < \kappa,$$

and unbounded, i.e.

$$\sup \{ f(\alpha) \mid \alpha < \lambda \} = \kappa.$$

The identity $(\alpha \mapsto \alpha)$ is a cofinal function on every limit ordinal, for example, but $(n \mapsto \aleph_n)$ is also cofinal, from ω to $\aleph_\omega$.

***x12.16.** Prove that for all von Neumann cardinals $\lambda \leq \kappa$, there exists a cofinal function $f : \lambda \rightarrow \kappa$ if and only if $cf(\lambda) = cf(\kappa)$. **HINT:** The cofinality operation $cf(\kappa)$ is defined in 9.23. The problem calls for the verification of several simple results of cardinal arithmetic about von Neumann cardinals, without assuming the full Axiom of Choice.

x12.17. For every regular λ , $cf(\aleph_\lambda) = \lambda$, so there exists cardinals of every regular cofinality.

***x12.18.** Prove that there exist singular von Neumann cardinals of every regular cofinality.

x12.19. For every von Neumann cardinal λ with $cf(\lambda) \geq \aleph_1$, the set $\mathcal{V}_\lambda$ is a Zermelo universe which further satisfies the following special case of the Replacement Axiom: if F is a definite operation such that $x \in \mathcal{V}_\lambda \implies F(x) \in \mathcal{V}_\lambda$, then the image $F[A]$ of every countable $A \in \mathcal{V}_\lambda$ is also in $\mathcal{V}_\lambda$.

12.28. Definition. (AC) An uncountable cardinal number κ is **strongly inaccessible** if it is regular and

$$\lambda < \kappa \implies 2^\lambda < \kappa.$$

***x12.20. (AC)** If κ is strongly inaccessible, then $\mathcal{V}_\kappa$ is a pure and grounded Z-F universe.

***x12.21. (AC)** If a pure and grounded set M is a Z-F universe, then $M = \mathcal{V}_\kappa$, for a strongly inaccessible cardinal κ .

12.29. Frege cardinals. We have followed Cantor in his approach to the theory of cardinal numbers, by which the property

$$A =_c |A| \tag{12.44}$$

is most fundamental. There is another approach due to Frege, which takes $|A|$ to be not a set of “units” equinumerous with A , but the abstract notion of “being equinumerous with A .” Frege understands “1”, for example, as the common property of all singletons. To model this idea in set theory, it is not important to define $|A|$ so that it is equinumerous with A , in fact, it is not even necessary for $|A|$ to be a set! The only important property of cardinals is the last one,

$$A =_c B \iff |A| = |B|, \tag{12.45}$$

which (in effect) makes the operation $|A|$ a “determining surjection” of the “equivalence condition” $=_c$, with the cardinal numbers as the “quotient class,” in the natural extension to classes of the terminology in **x4.5**. Frege tried to capture this idea by setting

$$|A| = \{X \mid X =_c A\}, \tag{12.46}$$

but the class $\{X \mid X =_c A\}$ is not a set (when $A \neq \emptyset$, easily) and the (necessary for the theory) assumption that it is led Frege to a contradiction.

Von Neumann cardinals reconcile the Cantor and Frege approaches by satisfying both (12.44) and (12.45), but their definition depends on both the Axioms of Choice and Replacement. Problem **x12.23** describes another approach, due to Scott, which succeeds in defining *Frege cardinals* without the Axiom of Choice, but (essentially) only for pure, grounded sets. Scott's construction is important, not so much for rescuing Frege cardinals (since little cardinal arithmetic can be done without **AC** anyway), but for the simplicity and elegance of the method, which has many uses beyond the present one. First we describe Scott's general method, and then its application to Frege cardinals.

12.30. Definition. *An equivalence condition on a class A is any binary, definite condition $\sim$ which has the properties of an equivalence relation, i.e. for all objects $x, y, z \in A$*

$$x \sim x, \quad x \sim y \implies y \sim x, \quad x \sim y \ \& \ y \sim z \implies x \sim z.$$

*A unary, definite operation F is **determining for** $\sim$ if*

$$x \sim y \iff F(x) = F(y) \quad (x, y \in A);$$

*the class of values of F for arguments in A is the **quotient class** of A by $\sim$ determined by F ,*

$$F[A] =_{\text{df}} \{F(x) \mid x \in A\}.$$

For example, the condition $=_o$ of similarity is an equivalence condition on the class of well ordered sets, and the von Neumann ordinal assignment $\text{ord}(U)$ is a determining operation for it, with quotient the class ON of ordinals. The condition $=_c$ of equinumerosity is an equivalence condition on the class of well orderable sets, and the von Neumann cardinal $|A|$ operation is determining for it, with quotient the class of von Neumann cardinals.

We can think of an equivalence condition $\sim$ on a class A very much as if A were a set and $\sim \subseteq A \times A$ an ordinary equivalence relation on it. There is no easy way to define a determining operation for $\sim$, however, because the classical construction of equivalence classes **4.14** truly leads to “classes” which need not be sets in this case: this is the problem with Frege's definition of the number 1 above.

x12.22. (Scott) Suppose $\sim$ is an equivalence condition on a class A of pure, grounded sets, and for each $x \in A$ let

$$\begin{aligned} \rho(x) &=_{\text{df}} (\mu\alpha \in ON)(\exists y \in \mathcal{V}_\alpha)[y \sim x], \\ F(x) &=_{\text{df}} \{y \in \mathcal{V}_{\rho(x)} \mid y \sim x\}. \end{aligned}$$

Prove that F is a determining operation for $\sim$ on A .

x12.23. (Scott) Define the **Scott cardinal** $|A|_s$ of every set A which is equinumerous with a pure, grounded set, so that for all such sets A and B ,

$$A =_c B \iff |A|_s = |B|_s.$$

Appendix A

THE REAL NUMBERS

In this Appendix we will show how the rational and the real numbers can be represented faithfully in set theory as the natural numbers are; that is, we will identify some characteristic, set theoretic properties of these systems and we will prove from the axioms of **ZDC** the existence and uniqueness (up to isomorphism) of structured sets with these properties. The proofs are quite simple as far as set theory goes, but they use ideas from algebra and analysis, which we will present in outline.

The basic tool we need is the construction of quotients of a set A by an equivalence relation on A described in Problem **x4.5**. Recall that a **determining surjection** for an equivalence relation $\sim$ on a set A is any surjection

$$\pi : A \rightarrow B$$

such that for all $x, y \in A$,

$$x \sim y \iff \pi(x) = \pi(y).$$

When this holds, we call B a **quotient of A by $\sim$** . The *canonical surjection* of $\sim$ is the mapping

$$x \mapsto [x/\sim] \quad (x \in A)$$

with quotient the set of equivalence classes $[A/\sim]$, but there may be others, more illuminating of the situation, e.g. those described in Problems **x4.6**, **x4.7** and **x4.8**. Determining surjections are especially useful in the study of congruence relations.

A.1. Definition. Suppose $\sim$ is an equivalence relation on A and

$$f : A \times A \rightarrow A$$

is a binary function. We call $\sim$ a **congruence for f** if for all $x, x', y, y' \in A$,

$$x \sim x' \ \& \ y \sim y' \implies f(x, y) \sim f(x', y').$$

The material in this Appendix can be read after Chapter 10. It assumes some theoretical knowledge of the Calculus, and it is not a prerequisite for understanding the remainder of these Notes.

Similarly, $\sim$ is a **congruence** for a binary relation $P \subseteq A \times A$, if for all x, x', y, y' ,

$$x \sim x' \ \& \ y \sim y' \implies [xPy \iff x'Py].$$

We can obviously define the notion of congruence for functions and relations of any number of arguments, in the same way.

The next theorem deals with one of the simplest and most basic algebraic constructions.

A.2. Theorem. *Let $\pi : A \rightarrow B$ be a determining surjection of some equivalence relation $\sim$, so that for all x, y in A , $x \sim y \iff \pi(x) = \pi(y)$.*

(1) *If $\sim$ is a congruence for a function $f : A \times A \rightarrow A$, then there exists exactly one function $f^\pi : B \times B \rightarrow B$ on the quotient B which satisfies the identity*

$$f^\pi(\pi(x), \pi(y)) = \pi(f(x, y)) \quad (x, y \in A). \quad (\text{A.1})$$

(2) *If $\sim$ is a congruence for a relation $P \subseteq A \times A$, then there exists exactly one relation $P^\pi \subseteq B \times B$ on the quotient B which satisfies the condition*

$$\pi(x)P^\pi\pi(y) \iff xPy \quad (x, y \in A).$$

Proof. The form of (A.1) makes it clear that at most one function can satisfy it, so it is enough to show that at least one function does. Put

$$f^\pi =_{\text{df}} \{((\pi(x), \pi(y)), \pi(z)) \mid x, y, z \in A \ \& \ f(x, y) = z\}.$$

To verify that the set of pairs f^π is a function, we must check that

$$((u, v), w), ((u, v), w') \in f^\pi \implies w = w'. \quad (\text{A.2})$$

From the hypothesis of (A.2) and the definition of f^π , there exist $x, y, z \in A$ such that

$$u = \pi(x), v = \pi(y), \quad w = \pi(z), f(x, y) = z$$

and also $x', y', z' \in A$, so that

$$u = \pi(x'), \quad v = \pi(y'), \quad w' = \pi(z'), \quad f(x', y') = z'.$$

It follows that

$$\pi(x) = \pi(x'), \quad \pi(y) = \pi(y'), \quad \pi(f(x, y)) = \pi(f(x', y'))$$

since $\sim$ is determined by π and it is a congruence for f , and the last of these equalities yields the desired $w = w'$.

The characteristic property (A.1) of f^π follows immediately from its definition and the proof of Part (2) is similar. $\dashv$

A.3. Exercise. *Prove Part (2) of A.2.*

The axiomatic characterizations of the rationals and the reals are based on the notion of an *ordered field*, which codifies the basic properties of addition, multiplication and ordering in these number systems.

A.4. Definition. *A field is a structured set*

$$(F, 0, 1, +, \cdot)$$

of objects with the following properties.

(F1) $0, 1 \in F$, $0 \neq 1$, and $+$, $\cdot$ are binary functions on F .

(F2) *The addition function $+$ satisfies the identities*

$$1. (x + y) + z = x + (y + z),$$

$$2. x + y = y + x,$$

$$3. x + 0 = x,$$

and for every x , there exists some x' , such that $x + x' = 0$.

(F3) *The multiplication function $\cdot$ satisfies the identities*

$$1. (x \cdot y) \cdot z = x \cdot (y \cdot z),$$

$$2. x \cdot y = y \cdot x,$$

$$3. x \cdot 1 = x,$$

and for every $x \neq 0$, there exists some x'' , such that $x \cdot x'' = 1$.

(F4) *Addition and multiplication together satisfy the identity*

$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

A.5. Lemma. *Every field F has the following properties:*

(1) *For each x there exists exactly one x' such that $x + x' = 0$, and we denote it $-x$; for every $x \neq 0$ there exists exactly one x'' such that $x \cdot x'' = 1$ and we denote it by x^{-1} .*

$$(2) x \cdot 0 = 0.$$

$$(3) x \cdot y = 0 \implies x = 0 \vee y = 0.$$

$$(4) (-x) \cdot y = -(x \cdot y).$$

Proof. (1) If $x + x' = 0$ and $x + y = 0$, then from the axioms

$$\begin{aligned} y = y + 0 = 0 + y &= (x + x') + y = x + (x' + y) \\ &= x + (y + x') = (x + y) + x' \\ &= 0 + x' = x' + 0 = x'. \end{aligned}$$

The proof about x^{-1} is similar.

(2) $x \cdot 0 = x \cdot (0 + 0) = x \cdot 0 + x \cdot 0$, and therefore

$$\begin{aligned} 0 = x \cdot 0 + -(x \cdot 0) &= (x \cdot 0 + x \cdot 0) + -(x \cdot 0) \\ &= x \cdot 0 + ((x \cdot 0) + -(x \cdot 0)) = x \cdot 0 + 0 = x \cdot 0. \end{aligned}$$

(3) If $y \neq 0$, then some y^{-1} exists such that $y \cdot y^{-1} = 1$, so that

$$x = x \cdot 1 = x \cdot (y \cdot y^{-1}) = (x \cdot y) \cdot y^{-1} = 0 \cdot y^{-1} = y^{-1} \cdot 0 = 0.$$

(4) $x \cdot y + (-x) \cdot y = y \cdot x + y \cdot (-x) = y \cdot (x + (-x)) = y \cdot 0 = 0$, and (1) implies that $(-x) \cdot y = -(x \cdot y)$. □

We gave this proof in full as an example, justifying each step from the field axioms. In the future we will cut corners, skip details or (more often) use identities which obviously hold in every field without proof or explicit mention.

A.6. Exercise. Every field F satisfies the identity

$$(x + y)^2 = x^2 + 2xy + y^2,$$

where $2 = 1 + 1$. (Give the proof in full detail.)

A.7. Exercise. The doubleton $\{0, 1\}$ of the first two natural numbers is a field, with the obvious operations, and in this field $1 + 1 = 0$. It follows that the field axioms do not imply $1 + 1 \neq 0$ and we must be a bit careful!

A.8. Definition. An ordered field is a structured set

$$(F, 0, 1, +, \cdot, \leq)$$

where $(F, 0, 1, +, \cdot)$ is a field, the binary relation $\leq$ is a linear ordering of F and the following conditions hold for all $x, y, z \in F$:

$$\begin{aligned} x \leq y &\implies x + z \leq y + z, \\ z > 0 \ \& \ x \leq y &\implies z \cdot x \leq z \cdot y, \end{aligned}$$

where $z > 0$ naturally abbreviates $0 \leq z$ & $z \neq 0$.

A.9. Exercise. *In every ordered field,*

$$z > 0 \ \& \ x < y \implies z \cdot x < z \cdot y.$$

A.10. Lemma. *Every element x in an ordered field F satisfies the inequality $x \cdot x = x^2 \geq 0$, so that $0 < 1$ and for all x , $x > 0 \implies x + 1 > 0$.*

Proof. If $x = 0$, then $x^2 = 0 \geq 0$, and if $x > 0$, then $x \cdot x \geq x \cdot 0 = 0$, so that the only interesting case is when $x < 0$. Adding $-x$ to both sides of this inequality, we get $0 < -x$, so that we can multiply $x < 0$ by $-x$ and we get $(-x) \cdot x < (-x) \cdot 0$, i.e. $-(x^2) < 0$ from the preceding Lemma, and adding x^2 to this inequality we get $0 < x^2$. The conclusion $0 < 1$ follows because $0 \neq 1$ and $1 = 1^2$, and the last claim holds because $0 < x \implies 1 < x + 1$, so that $0 < x + 1$ by the transitivity of $\leq$. $\dashv$

The Lemma makes it clear that we will not find in ordered fields the anomaly $1 + 1 = 0$ of Exercise A.7. Something much stronger is true.

A.11. Lemma. *Suppose F is an ordered field and set*

$$N_F = \bigcap \{X \subseteq F \mid 0 \in X \ \& \ (\forall x)[x \in X \implies x + 1 \in X]\};$$

*it follows that $(N_F, 0, (x \mapsto x + 1))$ is a system of natural numbers. The members of N_F are the **natural numbers of F** .*

Proof. By $(x \mapsto x + 1)$ we mean the function S which associates with each $x \in N_F$ the element $x + 1$ of F , which is also a member of N_F by the definition. The first three axioms of Peano are obvious and the fourth $(x + 1 \neq 0)$ holds because by the definition,

$$N_F \subseteq \{x \in F \mid 0 \leq x\},$$

and by the Lemma, $x \geq 0 \implies x + 1 \geq 1 > 0$. The Induction Axiom follows immediately from the definition of N_F as an intersection. $\dashv$

A.12. Exercise. *Suppose F is an ordered field, $N = N_F$ is the set of its natural numbers and $+_N, \cdot_N, \leq_N$ the addition, multiplication and the wellordering of N_F as these are defined in Chapter 5. Prove that these functions and the relation $\leq$ coincide with the respective objects in F , e.g.*

$$(\forall x, y \in N)[x +_N y = x + y].$$

The basic idea for the axiomatic characterization of the rationals is that they are an ordered field and that every fraction is a quotient of integers,

$$\frac{u - v}{m} = m^{-1} \cdot (u - v),$$

where m, u and v are natural numbers and $m \neq 0$. This simple observation yields not only the axioms for the rationals, but also proofs of their existence and uniqueness.

A.13. Definition. *A system of rational numbers is any ordered field F which satisfies the condition*

$$(\forall x)(\exists m, u, v \in N_F)[m \neq 0 \ \& \ x = m^{-1} \cdot (u - v)].$$

A.14. Theorem. Uniqueness of the rationals. *For any two systems of rational numbers F^1, F^2 there exists exactly one bijection*

$$\pi : F^1 \xrightarrow{\sim} F^2$$

which is an isomorphism, i.e.

1. $\pi(0^1) = 0^2, \pi(1^1) = 1^2$.
2. $\pi(x +^1 y) = \pi(x) +^2 \pi(y), \pi(x \cdot^1 y) = \pi(x) \cdot^2 \pi(y)$.
3. $x \leq^1 y \iff \pi(x) \leq^2 \pi(y)$.

In stating this theorem we decorated the various objects with the superscripts 0 or 1 to clarify the field to which they belong, e.g. $+^1$ is addition in F^1 and 0^2 is the zero element of F^2 . This is awkward and unnecessary, because it is always obvious which superscript is needed: e.g. the identity $\pi(0) = 0$ cannot mean anything else but $\pi(0^1) = 0^2$, since π is a function with domain F^1 and image F^2 . In the proof and in the future we will follow the general algebraic practice by which all the zero elements are 0, all additions are $+$, etc. We will also begin to skip the $\cdot$ of multiplication,

$$xy =_{\text{df}} x \cdot y.$$

Proof. By the uniqueness of the natural numbers and A.11, we know that there exists a “canonical” isomorphism

$$\rho : N^1 \xrightarrow{\sim} N^2,$$

where N^1 and N^2 are the sets of natural numbers in F^1 and F^2 , respectively. We set

$$\pi = \{(m^{-1}(u - v), \rho(m)^{-1}(\rho(u) - \rho(v))) \mid m, u, v \in N^1, m \neq 0\},$$

so that $\pi \subseteq F^1 \times F^2$ and it is enough to show first that π is a function, then that it is a bijection, and finally that it is an isomorphism, as we defined this in the formulation of the theorem.

To verify first that π is a function, we must show that if

$$m_1^{-1}(u_1 - v_1) = m_2^{-1}(u_2 - v_2), \quad (\text{A.3})$$

then

$$\rho(m_1)^{-1}(\rho(u_1) - \rho(v_1)) = \rho(m_2)^{-1}(\rho(u_2) - \rho(v_2)). \quad (\text{A.4})$$

The field axioms imply easily that (A.3) and (A.4) are respectively equivalent to

$$\begin{aligned} m_2 u_1 + m_1 v_2 &= m_1 u_2 + m_2 v_1, \\ \rho(m_2)\rho(u_1) + \rho(m_1)\rho(v_2) &= \rho(m_1)\rho(u_2) + \rho(m_2)\rho(v_1), \end{aligned}$$

and the first of these yields immediately

$$\rho(m_2 u_1 + m_1 v_2) = \rho(m_1 u_2 + m_2 v_1)$$

which in turn implies the second, because ρ is an isomorphism of N^1 with N^2 and it respects addition and multiplication by Problem x5.4.

The same simple method can be used directly to prove the additional conclusions, that π is one-to-one and finally an isomorphism. $\dashv$

A.15. Exercise. *Work out in detail the proofs of*

$$\begin{aligned} \pi(x + y) &= \pi(x) + \pi(y), \\ x \leq y &\iff \pi(x) \leq \pi(y). \end{aligned}$$

A.16. Theorem. Existence of the rationals. *There exists a system of rational numbers.*

Proof. If we have the rationals, we can define the set

$$A = \{(m, u, v) \mid m, u, v \in N \text{ \& } m \neq 0\}$$

of triples of integers, and on it the relation

$$(m, u, v) \sim (m', u', v') \iff_{\text{df}} m'u + mv' = mu' + m'v,$$

which (quite obviously) satisfies

$$(m, u, v) \sim (m', u', v') \iff \frac{u - v}{m} = \frac{u' - v'}{m'}.$$

This means that $\sim$ is an equivalence relation determined by the surjection

$$\pi : A \twoheadrightarrow Q, \quad \pi(m, u, v) = \frac{u - v}{m}. \quad (\text{A.5})$$

We do not have the rationals yet, but we have A and $\sim$: the idea for the proof is to define the rationals as a quotient of A by $\sim$ so that (A.5) holds. First we must show that

(1) $\sim$ is an equivalence relation. As an example, we verify that $\sim$ is transitive. From its definition, if

$$(m_1, u_1, v_1) \sim (m_2, u_2, v_2) \ \& \ (m_2, u_2, v_2) \sim (m_3, u_3, v_3),$$

then the identities

$$\begin{aligned} m_2 u_1 + m_1 v_2 &= m_1 u_2 + m_2 v_1, \\ m_3 u_2 + m_2 v_3 &= m_2 u_3 + m_3 v_2 \end{aligned}$$

hold in the natural numbers, and if we multiply the first of these by m_3 and the second by m_1 and then we add them, we get

$$\begin{aligned} m_3 m_2 u_1 + m_3 m_1 v_2 + m_1 m_3 u_2 + m_1 m_2 v_3 \\ = m_3 m_1 u_2 + m_3 m_2 v_1 + m_1 m_2 u_3 + m_1 m_3 v_2. \end{aligned}$$

Subtract now $m_3 m_1 v_2$ and $m_1 m_3 u_2$ from the two sides and divide by m_2 , which gives

$$m_3 u_1 + m_1 v_3 = m_1 u_3 + m_3 v_1,$$

i.e. $(m_1, u_1, v_1) \sim (m_3, u_3, v_3)$. Reflexivity and symmetry are proved in the same way.

(2) *Definition of the rationals.* Since $\sim$ is an equivalence relation, there exists a surjection

$$\pi : A \twoheadrightarrow Q$$

onto some set Q which determines it, so that

$$(m_1, u_1, v_1) \sim (m_2, u_2, v_2) \iff \pi(m_1, u_1, v_1) = \pi(m_2, u_2, v_2).$$

This Q is the set of rationals in the system under construction, and it remains only to specify 0 and 1, to define addition, multiplication and the ordering and finally to prove that the axioms for the rationals hold. To help follow the argument we will start right away using the notation

$$\frac{u - v}{m} =_{\text{df}} \pi(m, u, v)$$

as an abbreviation, i.e. without defining separately “subtraction” or “division.”

The zero and the one are defined in the obvious way,

$$0 = \frac{0 - 0}{1} =_{\text{df}} \pi(1, 0, 0), \quad 1 = \frac{1 - 0}{1} =_{\text{df}} \pi(1, 1, 0).$$

(3) *Addition of rationals.* With the representation of rationals as quotients of a difference of numbers by a number which we are using, the classical formula for addition of fractions takes the form

$$\frac{u_1 - v_1}{m_1} + \frac{u_2 - v_2}{m_2} = \frac{(m_2 u_1 + m_1 u_2) - (m_2 v_1 + m_1 v_2)}{m_1 m_2}.$$

So we define first on the set A the binary function f_+ which corresponds to this formula,

$$f_+((m_1, u_1, v_1), (m_2, u_2, v_2)) = (m_1 m_2, (m_2 u_1 + m_1 u_2), (m_2 v_1 + m_1 v_2)).$$

With a bit of arithmetic we can prove that for all $x, y, x', y' \in A$,

$$x \sim x' \ \& \ y \sim y' \implies f_+(x, y) \sim f_+(x', y'),$$

i.e. $\sim$ is a congruence for f_+ . It follows by **A.2** that there exists a (unique) function

$$+ : Q \times Q \rightarrow Q$$

which satisfies the identity

$$\pi(x) + \pi(y) = \pi(f_+(x, y)) \quad (\pi(x), \pi(y) \in Q).$$

Verification of the axioms (F2) for addition needs a bit more of arithmetic, but at least the condition for 0 is obvious:

$$\pi(m, u, v) + \pi(1, 0, 0) = \pi(m \cdot 1, 1 \cdot u + m \cdot 0, 1 \cdot v + m \cdot 0) = \pi(m, u, v).$$

(4) *Multiplication of rationals.* Following the same method, we define first the function $f : A \times A \rightarrow A$ which corresponds to multiplication when we represent rationals by triples of natural numbers,

$$f((m_1, u_1, v_1), (m_2, u_2, v_2)) =_{\text{df}} (m_1 m_2, u_1 u_2 + v_1 v_2, u_1 v_2 + u_2 v_1),$$

we verify next that $\sim$ is a congruence for f . and we define the multiplication operation on fractions $\cdot$ by **A.2** so that it satisfies the identity

$$\pi(x) \cdot \pi(y) = \pi(f(x, y)) \quad (\pi(x), \pi(y) \in Q).$$

Verification of axioms (F3) and (F4) requires just a few computations.

(5) *Ordering of the rationals.* The critical equivalence in this case is

$$\frac{u_1 - v_1}{m_1} \leq \frac{u_2 - v_2}{m_2} \iff m_1 v_2 + m_2 u_1 \leq m_2 v_1 + m_1 u_2.$$

We first define the relation $P \subseteq A \times A$ by

$$(m_1, u_1, v_1) P (m_2, u_2, v_2) \iff_{\text{df}} m_1 v_2 + m_2 u_1 \leq m_2 v_1 + m_1 u_2,$$

we verify that $\sim$ is a congruence for P and using **A.2** we define $\leq$ on the quotient Q so that

$$\pi(x) \leq \pi(y) \iff xPy \quad (\pi(x), \pi(y) \in Q).$$

That $\leq$ is a linear ordering and the structures set

$$(Q, 0, 1, +, \cdot, \leq)$$

is an ordered field follow with little difficulty.

It remains to verify that Q is a system of rational numbers.

Lemma 1. *For each natural number k , $\pi(1, k, 0) \in N_Q$, i.e. the rational $\pi(1, k, 0)$ belongs to the set of natural numbers of the ordered field Q .*

Proof. By induction on k , $\pi(1, 0, 0) = 0$ (by definition) and (easily) by the definition of rational addition

$$\pi(1, Sk, 0) = \pi(1, k, 0) + 1,$$

so that $\pi(1, k, 0) \in N_Q \implies \pi(1, Sk, 0) \in N_Q$.

Lemma 2. *For all $(m, u, v) \in A$,*

$$\pi(m, u, v) = \pi(1, m, 0)^{-1}(\pi(1, u, 0) - \pi(1, v, 0)), \quad (\text{A.6})$$

where $^{-1}$ and $-$ are the multiplicative and additive inverse (partial) functions of the field Q .

Proof. Having proved already that Q is a field, we know that (A.6) is equivalent to

$$\pi(1, m, 0)\pi(m, u, v) + \pi(1, v, 0) = \pi(1, u, 0),$$

and the latter identity is easy to verify with a direct computation.

The two Lemmas together show that the structured set $(Q, 0, 1, +, \cdot, \leq)$ satisfies the characteristic property of the rationals and this completes the proof. $\dashv$

As we did with the natural numbers, we now fix a specific system of rational numbers

$$(Q, 0, 1, +, \cdot, \leq)$$

whose elements we will henceforth call **rational**s. This is convenient, it helps avoid awkward expressions like “members of any system of rational numbers” and the like. However, it is important to emphasize (once more) that the significant mathematical fact is the existence and uniqueness up to isomorphism of one such system: it was precisely the corresponding mathematical facts about the natural numbers that we have used in the proofs of this Appendix, not the specific identity of “the natural numbers.”

A.17. Exercise. *The set Q of rationals is countable.*

A.18. Exercise. *In the proof of Part (1) of the theorem we “subtracted” the same number from an identity and then “divided” an identity by the same number. Justify these steps by verifying the following two properties of the natural numbers:*

$$\begin{aligned}x + y = x + z &\implies y = z, \\c \cdot x = c \cdot y \text{ \& } c \neq 0 &\implies x = y.\end{aligned}$$

A.19. Exercise. *For every ordered field F , there exists exactly one imbedding of the rationals in F , i.e. an injection*

$$\pi : Q \rightarrow F$$

which satisfies the identities

$$\begin{aligned}\pi(0) &= 0, \quad \pi(1) = 1, \\ \pi(x + y) &= \pi(x) + \pi(y), \quad \pi(xy) = \pi(x)\pi(y), \\ x \leq y &\iff \pi(x) \leq \pi(y).\end{aligned}$$

It follows that the image $\pi[Q] \subseteq F$ of π is a system of rational numbers (with the 0 and 1 of F and the restrictions of the operations and the ordering of F).

There is a beautiful theorem of Cantor which characterizes the ordering of the rationals independently of their algebraic structure. For it, we need first some definitions.

A.20. Definition. *Suppose $\leq$ is a linear ordering on a set A and $B \subseteq A$. We call B **dense in A** if*

$$(\forall x, y \in A)[x < y \implies (\exists b \in B)[x < b \text{ \& } b < y]].$$

*A linear ordering $\leq$ is **dense in itself** if its field (A) is dense in A .*

A.21. Exercise. *The ordering of every ordered field is dense in itself and has no minimum or maximum element.*

A.22. Theorem. (Cantor) *Every linear, dense in itself ordering $\leq_A$ without minimum or maximum element on a countable set A is **similar** with the ordering $\leq_Q$ of the rational numbers, i.e. there exists an order-preserving correspondence $f : Q \rightarrow A$.*

Proof. From the hypothesis and the fact that Q is countable, there exist enumerations without repetitions

$$Q = \{r_0, r_1, \dots\}, \quad A = \{a_0, a_1, \dots\}$$

of Q and A . We will define by recursion a sequence

$$f_0, f_1, \dots,$$

with the following properties, for every $n \in N$.

1. f_n is a *finite, partial function* from Q to A , i.e. $\text{Function}(f_n)$ & $f_n \subseteq Q \times A$, and f_n is finite as a set of ordered pairs.
2. f_n is monotone and one-to-one on its domain, i.e.

$$x, y \in \text{Domain}(f_n) \text{ \& } x <_Q y \implies f_n(x) <_A f_n(y).$$

3. $f_n \subseteq f_{n+1}$.
4. $\{r_0, r_1, \dots, r_n\} \subseteq \text{Domain}(f_n)$.
5. $\{a_0, a_1, \dots, a_n\} \subseteq \text{Image}(f_n)$.

If we can succeed in this, then the union $f =_{\text{df}} \bigcup_{n=0}^{n=\infty} f_n$ is (easily) a function by (3), it is one-to-one and monotone by (2) and

$$\text{Domain}(f) = Q, \quad \text{Image}(f) = A$$

by (4) and (5).

At the basis of the recursive definition we start with

$$p_0 = \{(r_0, a_0)\},$$

so that all the conditions of the result hold trivially.

Suppose now that we have already defined f_n and enumerate its finite domain of definition and image in increasing order:

$$\begin{aligned} D_n &= \{x_0 <_Q x_1 <_Q \dots <_Q x_m\}, \\ I_n &= \{y_0 <_A y_1 <_A \dots <_A y_m\}. \end{aligned}$$

Since f_n is monotone, we have

$$f_n(x_i) = y_i \quad (i = 0, \dots, m).$$

We construct the next f_{n+1} in two steps, i.e. first we will define some $f'_{n+1} \supseteq f_n$ which satisfies (1) - (4) and then $f_{n+1} \supseteq f'_{n+1}$ which satisfies all (1) - (5).

STEP 1. If $r_{n+1} \in \text{Domain}(f_n)$, set $f'_{n+1} = f_n$. Otherwise there are three cases.

CASE 1. $r_{n+1} <_Q x_0$. In this case we find some $y' \in A$ satisfying $y' <_A y_0$ (which exists because A has no minimum) and set

$$f'_{n+1} = f_n \cup \{(r_{n+1}, y')\}.$$

CASE 2. $r_{n+1} >_Q x_m$. In this case we find some $y' \in A$ satisfying $y' >_A y_m$ (which exists because A has no maximum) and we set

$$f'_{n+1} = f_n \cup \{(r_{n+1}, y')\}.$$

CASE 3. For some i , $x_i <_Q r_{n+1} <_Q x_{i+1}$. In this case we find some $y' \in A$ satisfying $y_i <_A y' <_A y_{i+1}$ (which exists because A is dense in itself) and we set

$$f'_{n+1} = f_n \cup \{(r_{n+1}, y')\}.$$

In all cases, the proof that f'_{n+1} satisfies (1) - (4) is simple.

In STEP 2 of the construction we consider the element a_{n+1} of A and we distinguish again cases: first if $a_{n+1} \in \text{Image}(f'_{n+1})$ (in which case we set $f_{n+1} = f'_{n+1}$) and if not, then three cases again, in symmetry with STEP 1. We skip the details. $\dashv$

The fundamental intuition about the real numbers is that on the one hand they are an ordered field, so that their arithmetic and ordering satisfy the same laws as the rationals, and on the other, they are in one-to-one correspondence with the points of the “complete” geometric line so that there are no “gaps” between them. In formulating the property of completeness we follow Dedekind.

A.23. Definition. A linear ordering $\leq$ on a set A is **complete** if every non-empty subset of A which has an upper bound has a least upper bound.

A system of real numbers is any **complete, ordered field**, i.e. any ordered field in which the ordering is complete.

A.24. Exercise. The ordering of the rationals is not complete, because the set

$$X = \{r \mid r^2 < 2\}$$

is bounded from above but has no least upper bound.

A.25. Lemma. Every complete, ordered field F has the **archimedean property**

$$(\forall x \in F)(\exists n \in N)[x < n],$$

i.e. the set $N = N_F$ of its natural numbers is not bounded from above.

Proof. Assume towards a contradiction that the set N has an upper bound, so that it has a least upper bound $x = \sup N$ by the completeness property. The element $x - 1$ is not an upper bound of N because $x - 1 < x$, so there must exist some $n \in N$, $x - 1 < n$: but this implies $x < n + 1$ which contradicts the assumption that x is an upper bound of N . $\neg$

A.26. Exercise. In every complete, ordered field F ,

$$\epsilon > 0 \implies (\exists n \in N) \left[\frac{1}{n+1} < \epsilon \right].$$

A.27. Exercise. In every complete, ordered field F ,

$$x < y \implies (\exists r \in Q)[x < r \text{ \& } r < y],$$

where $Q = Q_F$ is the set of rationals in F . (Density of the rationals.)

We now aim to show that *there exists a complete ordered field* and that *any two complete, ordered fields are isomorphic*. Because the completeness property is geometric (or topological), these proofs of existence and uniqueness depend on geometric ideas. Specifically, we will need some basic definitions and results from the *theory of limits* which is studied in Calculus. We will review these here, briefly but without limiting ourselves to the absolute minimum list of theorems necessary to prove the existence and uniqueness of the reals: we have included several Lemmas and Exercises because they support the proposition that the notion of a complete, ordered field represents faithfully our geometric intuitions about the real numbers.

Since we will be manipulating (infinite) sequences a great deal, we will use consistently the familiar, simple notation which avoids the introduction of a distinct name for each sequence and treats the variable as an index, so that a sequence ($n \mapsto a_n$) is denoted by $\langle a_n \rangle$ or $\langle a_0, a_1, \dots \rangle$. For example, the sequence

$$\langle \frac{1}{n+1} \rangle = \langle \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \dots \rangle$$

is the function $f : N \rightarrow Q$ defined by the formula

$$f(n) = \frac{1}{n+1}.$$

The **absolute value** function is defined in every ordered field in the usual way,

$$|x| =_{\text{df}} \max\{x, -x\} = \begin{cases} x, & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

A.28. Definition. Suppose $(F, 0, 1, +, \cdot, \leq)$ is an ordered field.

(1) A sequence $\langle x_n \rangle$ of elements of F **converges** to $x \in F$ or has **limit** x , if

$$(\forall \epsilon \in F, \epsilon > 0)(\exists K \in N)(\forall n \in N)[n \geq K \implies |x - x_n| < \epsilon].$$

We will use the notation

$$x_n \rightarrow x \iff_{\text{df}} \langle x_n \rangle \text{ converges to } x.$$

(2) A sequence $\langle x_n \rangle$ has the property of **Cauchy**, or (simply) is **Cauchy** if

$$(\forall \epsilon \in F, \epsilon > 0)(\exists K \in N)(\forall n, m \in N)[n, m \geq K \implies |x_n - x_m| < \epsilon].$$

A.29. Definition. For all $a < b$ in an ordered field F , the set

$$(a, b) =_{\text{df}} \{x \in F \mid a < x < b\} \quad (\text{A.7})$$

is the **open interval** with endpoints a and b . A set $G \subseteq F$ is **open** if it is a (possibly empty) union of open intervals, equivalently

$$x \in G \implies (\exists a < b)[x \in (a, b) \subseteq G].$$

We will also use the standard notations for closed and half-open intervals, e.g.

$$(a, b] = \{x \in F \mid a < x \leq b\}.$$

A.30. Exercise. Prove that the family of open sets in an ordered field is a topology and that the definition of limits for sequences in **A.28** is equivalent to the topological definition of limits given in **10.36**.

These definitions are notorious for the difficulty of understanding what they mean and learning how to use them. We emphasize that here we study them in the context of an arbitrary ordered field which need not be complete, for example, the rationals. It is useful to formulate conditions equivalent to convergence and the property of Cauchy, on the basis of the following notion.

A.31. Definition. A sequence $\langle x_n \rangle$ in an ordered field **settles** in an open interval (a, b) , if after a certain stage all its terms belong to some closed subinterval $[a', b'] \subseteq (a, b)$:

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff_{\text{df}} (\exists K, a', b')(\forall n \geq K)[a < a' \leq x_n \leq b' < b].$$

Notice that if $\langle x_n \rangle \rightsquigarrow (a, b)$, then all its terms after a certain stage belong to (a, b) ,

$$\langle x_n \rangle \rightsquigarrow (a, b) \implies (\exists K)(\forall n \geq K)[a < x_n < b];$$

this weaker property is all we need for many applications of the definition of $\langle x_n \rangle \rightsquigarrow (a, b)$.

A.32. Exercise. $\langle x_n \rangle \rightsquigarrow (a, b)$ if and only if there exists some $\delta > 0$ such that $a + \delta < b - \delta$ and the set $\{n \in \mathbb{N} \mid x_n \notin [a + \delta, b - \delta]\}$ is finite.

A.33. Exercise. For all open intervals I, J , $\langle x_n \rangle \rightsquigarrow I$ & $I \subseteq J \implies \langle x_n \rangle \rightsquigarrow J$.

A.34. Exercise. For all open intervals I, J ,

$$\langle x_n \rangle \rightsquigarrow I \text{ \& } \langle x_n \rangle \rightsquigarrow J \implies \langle x_n \rangle \rightsquigarrow I \cap J \implies I \cap J \neq \emptyset.$$

A.35. Exercise. Every sequence $\langle x_n \rangle$ which settles in some open interval (a, b) is **bounded**, i.e.

$$(\exists w)(\forall n)[x_n \leq w].$$

The next Lemma makes it possible in many cases to avoid the so-called “method of epsilonics,” which is illustrated by its proof.

A.36. Lemma. For every sequence in any ordered field F :

(1) $\langle x_n \rangle$ converges to x if and only if $\langle x_n \rangle$ settles in every open interval which contains x ,

$$x_n \rightarrow x \iff (\forall a, b \in F)[a < x < b \implies \langle x_n \rangle \rightsquigarrow (a, b)].$$

(2) $\langle x_n \rangle$ is Cauchy if and only if for every $\epsilon > 0$, there exists an open interval (a, b) in which $\langle x_n \rangle$ settles and such that $(b - a) \leq \epsilon$:

$$\langle x_n \rangle \text{ is Cauchy} \iff (\forall \epsilon > 0)(\exists a, b)[a < b \leq a + \epsilon \text{ \& } \langle x_n \rangle \rightsquigarrow (a, b)]$$

Proof. (1) If $x_n \rightarrow x$ and $a < x < b$, then the definition of convergence with

$$\epsilon = \frac{\min(x - a, b - x)}{2}$$

supplies a number K such that

$$n \geq K \implies |x - x_n| < \frac{\min(x - a, b - x)}{2},$$

which with a bit of inequality massaging implies that

$$n \geq K \implies a < a' = \frac{a + x}{2} < x_n < \frac{x + b}{2} = b' < b,$$

so that $\langle x_n \rangle \rightsquigarrow (a, b)$. For the other direction, for every $\epsilon > 0$, $\langle x_n \rangle \rightsquigarrow (x - \epsilon, x + \epsilon)$, so that for some K ,

$$n \geq K \implies |x - x_n| < \epsilon.$$

(2) If $\langle x_n \rangle$ is Cauchy, then for every $\epsilon > 0$ there exists some K such that $n, m \geq K \implies |x_n - x_m| < \frac{\epsilon}{4}$, which immediately implies that $\langle x_n \rangle \rightsquigarrow (x_K - \frac{\epsilon}{2}, x_K + \frac{\epsilon}{2})$, and this interval has length ϵ . In the other direction, for every $\epsilon > 0$ there exists some (a, b) with $(b - a) \leq \epsilon$, so that $\langle x_n \rangle \rightsquigarrow (a, b)$, and hence for some K ,

$$n, m \geq K \implies [a < x_n < b \ \& \ a < x_m < b \implies |x_n - x_m| < b - a \leq \epsilon],$$

which means that $\langle x_n \rangle$ is Cauchy. $\dashv$

A.37. Corollary. *If $\langle x_n \rangle$ converges to some x , then it is Cauchy.*

Proof. For every $\epsilon > 0$, $\langle x_n \rangle \rightsquigarrow (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2})$ by (1) of **A.36**, so it is Cauchy by (2) of **A.36**. $\dashv$

A.38. Exercise. $x_n \rightarrow x \ \& \ x_n \rightarrow y \implies x = y$. This allows us to introduce the classical notation

$$x = \lim_n x_n \iff_{\text{df}} x_n \rightarrow x.$$

A.39. Lemma. *If $\langle x_n \rangle$ is Cauchy in a complete, ordered field F , then $\langle x_n \rangle$ has a limit, i.e. $x_n \rightarrow x$ with some x .*

Proof. Let

$$X =_{\text{df}} \{u \in F \mid (\exists v)[u < v \ \& \ \langle x_n \rangle \rightsquigarrow (u, v)]\}.$$

Since $\langle x_n \rangle$ is Cauchy, there exists some (c, d) such that $\langle x_n \rangle \rightsquigarrow (c, d)$, and hence

$$u \in X \implies u \leq d,$$

since $d < u \implies \langle x_n \rangle \rightsquigarrow (c, d) \cap (u, v) = \emptyset$ which is not possible; hence X is bounded from above and it must have a least upper bound

$$x = \sup X.$$

We will show that

$$a < x < b \implies \langle x_n \rangle \rightsquigarrow (a, b)$$

which implies $x_n \rightarrow x$ by **A.36**. Using the hypothesis $a < x < b$ and the fact that $\langle x_n \rangle$ is Cauchy, we find first some (u, v) such that

$$v - u < \min(x - a, b - x), \quad \langle x_n \rangle \rightsquigarrow (u, v). \quad (\text{A.8})$$

By definition, $u \in X$, and hence (a) $u \leq x$, because x is an upper bound of X . On the other hand, v is also an upper bound of X (because the assumption $\langle x_n \rangle \rightsquigarrow (u', v')$ with $v < u'$ implies that $\langle x_n \rangle$ settles in two disjoint intervals), and hence (b) $x \leq v$, since x is the least upper bound of X . Now (a) and (b) together yield (c) $u \leq x \leq v$, which together with (A.8) implies $a < u \leq x \leq v < b$, so that $\langle x_n \rangle \rightsquigarrow (a, b)$. $\dashv$

The next two basic theorems relate completeness as we defined it (following Dedekind) with the notion of completeness historically associated with the name of Cantor.

A.40. Theorem. The Nested Interval Property. *Suppose that every Cauchy sequence in an ordered field F converges, and that*

$$[x_0, y_0] \supseteq [x_1, y_1] \supseteq \cdots \quad (\text{A.9})$$

is a nested sequence of closed intervals such that

$$\lim_n (y_n - x_n) = 0; \quad (\text{A.10})$$

it follows that the intersection $\bigcap_n [x_n, y_n]$ is a singleton

$$\bigcap_n [x_n, y_n] = \{w\},$$

and its only member is the common limit of the sequences $\langle x_n \rangle$ and $\langle y_n \rangle$,

$$w = \lim_n x_n = \lim_n y_n.$$

Proof. The basic observation is that for every number K and every $\delta > 0$, $\langle x_n \rangle \rightsquigarrow [x_K - \delta, y_K + \delta]$ by (A.9). Now (A.10) implies that $\langle x_n \rangle$ is Cauchy and hence $x_n \rightarrow x$ for some x , using the hypothesis. In addition, $x < x_K \implies \langle x_n \rangle \rightsquigarrow (x - 1, x_K)$ by **A.36**, which contradicts the basic properties of the relation $\rightsquigarrow$ since $(x - 1, x_K) \cap [x_K, y_K] = \emptyset$, hence, $x_K \leq x$, for every K . By a similar argument $x \leq y_K$, for every K , so that in the end $x \in \bigcap_n [x_n, y_n]$. Symmetrically, $\langle y_n \rangle$ converges, $y = \lim_n y_n \in \bigcap_n [x_n, y_n]$, and for every n , $|x - y| \leq (y_n - x_n)$ which implies $x = y$ by (A.10) and completes the proof. $\dashv$

A.41. Theorem. *An ordered field F is complete if and only if it has the archimedean property (A.25) and every Cauchy sequence in F has a limit.*

Proof. One direction is known from Lemmas **A.25** and **A.39**, so it is enough to show that if F has the archimedean property and every Cauchy sequence in F converges, then F is complete.

Suppose then that X is a non-empty, bounded from above set in F , so that there exists some point $x_0 \in X$ and some upper bound y_0 of X . Beginning with $[x_0, y_0]$, we define by recursion a sequence of closed intervals $[x_n, y_n]$ which satisfy the following conditions:

1. $x_n \leq x_{n+1} < y_{n+1} \leq y_n$,
2. $(y_n - x_n) = 2^{-n}(y_0 - x_0)$,
3. $[x_n, y_n] \cap X \neq \emptyset$,
4. $(\forall x \in X)[x \leq y_n]$.

In detail, to define $[x_{n+1}, y_{n+1}]$ we distinguish two cases: if $w = \frac{1}{2}(x_n + y_n)$ is an upper bound of X , we set $[x_{n+1}, y_{n+1}] = [x_n, w]$, otherwise $[x_{n+1}, y_{n+1}] = [w, y_{n+1}]$. Proof that $[x_{n+1}, y_{n+1}]$ satisfies (1) - (4) is trivial.

Lemma. $\lim_n (y_n - x_n) = 0$.

Proof. The archimedean property implies that for every $\epsilon > 0$, there exists some natural number $K > 0$ such that

$$\frac{y_0 - x_0}{\epsilon} < K \leq 2^K,$$

where the inequality $K \leq 2^K$ is verified easily (by induction!). It follows that for every $n \geq K$,

$$(y_n - x_n) = 2^{-n}(y_0 - x_0) \leq 2^{-K}(y_0 - x_0) < \epsilon,$$

which completes the proof of the Lemma.

Now **A.40** implies that

$$\bigcap [x_n, y_n] = \{w\}$$

where $w = \lim_n x_n = \lim_n y_n$, and it is enough to verify that this common limit w is the least upper bound of X . We compute:

$$\begin{aligned} w < t &\implies \langle y_n \rangle \rightsquigarrow (w - 1, t) && \text{because } \lim_n y_n = w, \\ &\implies y_n < t && \text{for some } n, \\ &\implies t \notin X && \text{because } y_n \text{ is} \\ & && \text{an upper bound of } X, \end{aligned}$$

so that w is an upper bound of X . Also,

$$\begin{aligned} t < w &\implies \langle x_n \rangle \rightsquigarrow (t, w + 1) && \text{because } \lim_n x_n = w, \\ &\implies t < x_n && \text{for some } n, \\ &\implies (\exists x \in X)[t < x] && \text{by the definition of } x_n, \end{aligned}$$

so that there is no upper bound of X smaller than w . $\dashv$

It is worth pointing out here that there exist *Cauchy complete* ordered fields which are not archimedean, Problem ***xA.2**.

Following Cantor (up to a point), we will construct a complete, ordered field as a quotient of the Cauchy sequences on the rationals by the following, natural equivalence relation.

A.42. Definition. We call two sequences of rationals $\langle x_n \rangle$ and $\langle y_n \rangle$ **asymptotically equivalent** if their difference converges to 0, in symbols,

$$\langle x_n \rangle \approx \langle y_n \rangle \iff_{\text{df}} (x_n - y_n) \rightarrow 0.$$

A.43. Theorem. (1) Two Cauchy sequences $\langle x_n \rangle$ and $\langle y_n \rangle$ are asymptotically equivalent if and only if they settle in the same open intervals:

$$\langle x_n \rangle \approx \langle y_n \rangle \iff (\forall a < b)[\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle y_n \rangle \rightsquigarrow (a, b)].$$

(2) If $\langle x_n \rangle$ and $\langle y_n \rangle$ are Cauchy, then

$$\begin{aligned} \langle x_n \rangle \not\approx \langle y_n \rangle &\implies (\text{there exist open intervals } I, J) \\ &[I \cap J = \emptyset \ \& \ \langle x_n \rangle \rightsquigarrow I \ \& \ \langle y_n \rangle \rightsquigarrow J]. \end{aligned}$$

(3) The relation $\approx$ is an equivalence relation on the set

$$\mathcal{C}(F) =_{\text{df}} \{ \langle x_n \rangle \mid \langle x_n \rangle \text{ is Cauchy on } F \}.$$

Proof. (1) Suppose first that $\langle x_n \rangle \approx \langle y_n \rangle$ and $\langle x_n \rangle \rightsquigarrow (a, b)$, so that for some K_0 and some $\delta > 0$,

$$n \geq K_0 \implies a + \delta \leq x_n \leq b - \delta.$$

Using $\langle x_n \rangle \approx \langle y_n \rangle$, choose K_1 such that

$$\begin{aligned} n \geq K_1 &\implies |x_n - y_n| < \frac{\delta}{2} \\ &\implies x_n + \frac{\delta}{2} < y_n < x_n + \frac{\delta}{2}. \end{aligned}$$

From these two implications we get easily that

$$n \geq \max(K_0, K_1) \implies a + \frac{\delta}{2} \leq y_n \leq b - \frac{\delta}{2},$$

so that $\langle y_n \rangle \rightsquigarrow (a, b)$. In the other direction, for every $\epsilon > 0$ there exists an open interval (a, b) with $b - a \leq \epsilon$ such that $\langle x_n \rangle \rightsquigarrow (a, b)$ by (2) of **A.36**, so by the hypothesis, we also have

$$n \geq K \implies [a < x_n < b \ \& \ a < y_n < b] \implies |y_n - x_n| < \epsilon.$$

(2) Directly from the definition of convergence,

$$\neg[(x_n - y_n) \rightarrow 0] \implies (\exists \epsilon > 0)(\forall K)(\exists n, m \geq K)[|x_n - y_m| \geq \epsilon]. \quad (\text{A.11})$$

By **A.36**, there exist open intervals I and J of length $< \frac{\epsilon}{2}$ (with this ϵ) such that $\langle x_n \rangle \rightsquigarrow I$ and $\langle y_n \rangle \rightsquigarrow J$. If some $z \in I \cap J$ existed, then for K sufficiently large so that

$$n, m \geq K \implies x_n \in I \ \& \ y_m \in J,$$

we would have

$$n, m \geq K \implies |x_n - y_m| \leq |x_n - z| + |z - y_m| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which contradicts (A.11).

(3) The reflexivity and symmetry of $\approx$ are trivial. To show its transitivity, notice that by hypothesis and (3) of Lemma A.36, for every (a, b) ,

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle y_n \rangle \rightsquigarrow (a, b) \iff \langle z_n \rangle \rightsquigarrow (a, b),$$

so that by (1), $\langle x_n \rangle \approx \langle z_n \rangle$. ⊢

A.44. Exercise. If $\langle x_n \rangle$ and $\langle y_n \rangle$ are both Cauchy and $x_n \rightarrow x$, then

$$\langle x_n \rangle \approx \langle y_n \rangle \iff y_n \rightarrow x.$$

At this point we could appeal to the existence of *some* quotient B of the set

$$\mathcal{C} = \mathcal{C}(Q) =_{\text{df}} \{ \langle r_n \rangle \mid \langle r_n \rangle \text{ is Cauchy in the field of rationals} \}$$

by $\approx$ and define the necessary functions and an ordering on B so that it becomes a complete, ordered field. This is one of the classical proofs of the existence of the real numbers, connected with the name of Cantor. Instead of this, we will construct *a specific quotient* of $\mathcal{C}$ by $\approx$ which simplifies the proof a bit and (more significantly) relates this construction with the other classical proof of the existence of the reals, following Dedekind. The basic idea of Dedekind was that a real number x is completely determined by (and hence can be “identified” with) the set

$$(-\infty, x) \cap Q =_{\text{df}} \{ r \in Q \mid r < x \}$$

of all rationals preceding it, and that the sets of the form $(-\infty, x) \cap Q$ can be characterized directly by three simple conditions.

A.45. Definition. A **Dedekind cut** is any set X of rational numbers which satisfies the following three conditions:

1. $X \neq \emptyset$, $(Q \setminus X) \neq \emptyset$.
2. $r < q$ & $q \in X \implies r \in X$.
3. $q \in X \implies (\exists r)[q < r \text{ & } r \in X]$.

We set

$$\mathcal{D} =_{\text{df}} \{ X \subseteq Q \mid X \text{ is a Dedekind cut} \}.$$

A.46. Exercise. A set $X \subseteq Q$ is a Dedekind cut if and only if it is non-empty, bounded from above, with no largest member and “downward closed,” i.e. such that $r < q$ & $q \in X \implies r \in X$.

The next theorem is basic for the proof of existence of the real numbers.

A.47. Theorem. For each Cauchy sequence of rationals $\langle x_n \rangle$, let

$$\pi(\langle x_n \rangle) =_{\text{df}} \{a \in Q \mid (\exists b)[a < b \text{ \& } \langle x_n \rangle \rightsquigarrow (a, b)]\};$$

it follows that each value $\pi(\langle x_n \rangle)$ is a Dedekind cut and that the function

$$\pi : \mathcal{C} \longrightarrow \mathcal{D}$$

is a surjection which determines the equivalence relation $\approx$, so that $\mathcal{D}$ is a quotient of $\mathcal{C}$.

Proof. That each $\pi(\langle x_n \rangle)$ is a Dedekind cut is quite easy from the definitions and the general properties of the relation $\rightsquigarrow$, and the equivalence

$$\langle x_n \rangle \approx \langle y_n \rangle \iff \pi(\langle x_n \rangle) = \pi(\langle y_n \rangle)$$

is an immediate corollary of **A.43**. The only thing which is not completely obvious is that for every cut X there exists a Cauchy sequence $\langle x_n \rangle \in \mathcal{C}$ in the rationals such that $\pi(\langle x_n \rangle) = X$. For this we construct a nested sequence of closed intervals *in the rationals*

$$[x_0, y_0] \supseteq [x_1, y_1] \supseteq \cdots$$

exactly as in the proof of **A.41**, beginning with some $x_0 \in X$ and some $y_0 \notin X$, so that, in fact, y_0 is an upper bound of X . We argue as in **A.41** that the non-decreasing sequence $\langle x_n \rangle$ is Cauchy, and that, in addition, for all n ,

$$x_n \in X \text{ \& } y_n \notin X,$$

because X is downward closed and has no largest member. Then we compute:

$$\begin{aligned} a \in \pi(\langle x_n \rangle) &\implies (\exists b)[a < b \text{ \& } \langle x_n \rangle \rightsquigarrow (a, b)] \\ &\implies (\exists n)[a < x_n] \\ &\implies a \in X \text{ because } x_n \in X. \end{aligned}$$

To see that $X \subseteq \pi(\langle x_n \rangle)$, notice that if $a \in X$, then there must exist some natural number K such that $a < x_K$, because the opposite supposition $(\forall n)[x_n \leq a]$ implies (easily) that a is the largest point in X , and X does not have a largest point. Thus, for $n \geq K$, $x_K \leq x_n \leq y_K$, hence, $\langle x_n \rangle \rightsquigarrow (x_K, y_K + 1)$ and $a \in \pi(\langle x_n \rangle)$. $\dashv$

A.48. Theorem. Existence of the real numbers. *There exists a complete, ordered field.*

Proof. For the domain of the field we take the set $\mathcal{D}$ of Dedekind cuts and for 0 and 1 we take the obvious:

$$\begin{aligned} 0 &=_{\text{df}} \{r \in \mathbb{Q} \mid r < 0\} = \pi(<0>), \\ 1 &=_{\text{df}} \{r \in \mathbb{Q} \mid r < 1\} = \pi(<1>). \end{aligned}$$

In order to define addition and multiplication on $\mathcal{D}$ we need the following two Lemmas, where all the sequences are Cauchy in the rationals.

$$\begin{aligned} <x_n> \approx <x'_n> \quad \& \quad <y_n> \approx <y'_n> \\ &\implies <x_n + y_n> \approx <x'_n + y'_n> \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} <x_n> \approx <x'_n> \quad \& \quad <y_n> \approx <y'_n> \\ &\implies <x_n \cdot y_n> \approx <x'_n \cdot y'_n> \end{aligned} \quad (\text{A.13})$$

These are the useful interpretations of the classical theorems from the theory of limits,

$$\begin{aligned} \lim_n (x_n + y_n) &= \lim_n x_n + \lim_n y_n, \\ \lim_n (x_n \cdot y_n) &= \lim_n x_n \cdot \lim_n y_n, \end{aligned}$$

in the case at hand, when the limits need not exist since the sequences are in the incomplete ordered field of the rationals. They are not hard to verify after all the preparatory work we have done and we will skip the details. The equivalences (A.12) and (A.13) assert that $\approx$ is a congruence in $\mathcal{C}$ for the functions

$$\begin{aligned} (<x_n>, <y_n>) &\mapsto <x_n + y_n>, \\ (<x_n>, <y_n>) &\mapsto <x_n \cdot y_n>, \end{aligned}$$

so that by **A.2** there exist functions $+$ and $\cdot$ on the quotient $\mathcal{D}$ which satisfy the identities

$$\begin{aligned} \pi(<x_n>) + \pi(<y_n>) &= \pi(<x_n + y_n>), \\ \pi(<x_n>) \cdot \pi(<y_n>) &= \pi(<x_n \cdot y_n>). \end{aligned}$$

We take these $+$ and $\cdot$ for the operations of addition and multiplication in $\mathcal{D}$.

Next we must show that $(\mathcal{D}, 0, 1, +, \cdot)$ is a field, but this part of the proof is quite trivial, if a bit tiring in its details (which we will skip). The existence of additive inverses, for example, follows from the obvious

$$\pi(<x_n>) + \pi(<-x_n>) = \pi(<x_n + (-x_n)>) = \pi(<0>) = 0,$$

where the only “delicate” point is the observation that if $\langle x_n \rangle$ is Cauchy, then so is $\langle -x_n \rangle$. To check the corresponding property for multiplication, given a Cauchy sequence $\langle x_n \rangle \not\approx \langle 0 \rangle$, set

$$y_n =_{\text{df}} \begin{cases} 1/x_n, & \text{if } x_n \neq 0, \\ 1, & \text{if } x_n = 0, \end{cases} \quad (\text{A.14})$$

verify that $\langle y_n \rangle$ is also Cauchy and then compute:

$$\pi(\langle x_n \rangle) \cdot \pi(\langle y_n \rangle) = \pi(\langle x_n \cdot y_n \rangle) = \pi(\langle 1 \rangle) = 1.$$

The basic observation (from (2) of **A.43**) is that

$$\begin{aligned} \langle x_n \rangle \not\approx \langle 0 \rangle &\implies (\exists \delta > 0)[\langle x_n \rangle \not\prec (-\delta, \delta)] \\ &\implies (\exists \delta > 0, K)(\forall n \geq K)|x_n| \geq \delta \end{aligned}$$

with which we begin the proof that $\langle y_n \rangle$ is Cauchy, but some epsilonics are unavoidable. The related result from the theory of limits is the assertion

$$\lim_n x_n \neq 0 \implies \lim_n \left(\frac{1}{x_n} \right) = \frac{1}{\lim_n x_n},$$

traditionally known as the first hard theorem in Calculus, when it is taught rigorously.

Next we define on $\mathcal{D}$ the relation

$$X \leq Y \iff_{\text{df}} X \subseteq Y,$$

which is certainly a partial ordering; it is also linear, because for any two Dedekind cuts X and Y , directly from the definition,

$$\begin{aligned} r \in (Y \setminus X) &\implies (\forall q \in X)[q < r] \ \& \ (\forall q < r)[q \in Y] \\ &\implies X \subseteq Y, \end{aligned}$$

and, of course, $X \neq Y \implies (\exists r)[r \in (X \setminus Y)] \vee [r \in (Y \setminus X)]$. Appealing once more to the definition of Dedekind cuts, we can also show easily that

$$I \subseteq \mathcal{D} \implies \bigcup I = Q \vee \bigcup I \in \mathcal{D}. \quad (\text{A.15})$$

(For example, if r were the largest point in the union $\bigcup I$, then $r \in X$ for some $X \in I$ and then r would also be the largest point $X \subseteq I$, which has no largest member.) From (A.15) we infer that every set $I \subseteq \mathcal{D}$ which is bounded above has a least upper bound, because

$$(\forall X \in I)[X \leq Z] \implies \bigcup I \subseteq Z \subsetneq Q \implies \bigcup I \in \mathcal{D},$$

and the union $\bigcup I$ is obviously the least upper bound of I in the relation $\leq = \subseteq$.

It remains to verify that for all $X, Y, Z \in \mathcal{D}$

$$X < Y \implies X + Z < Y + Z, \quad (\text{A.16})$$

$$[Z > 0 \text{ \& } X < Y] \implies Z \cdot X < Z \cdot Y, \quad (\text{A.17})$$

since these implications imply then immediately their versions with $\leq$. Considering the more difficult (A.17) as an example, choose first Cauchy sequences such that

$$\pi(\langle z_n \rangle) = Z, \quad \pi(\langle x_n \rangle) = X, \quad \pi(\langle y_n \rangle) = Y,$$

and verify (easily) from the definitions (and **A.43**) that there exist rationals

$$x^0 < x^1 < y^0 < y^1$$

satisfying

$$\langle x_n \rangle \rightsquigarrow (x^0, x^1), \quad \langle y_n \rangle \rightsquigarrow (y^0, y^1),$$

and that for each $\epsilon > 0$ we can find some z^0 and z^1 such that

$$0 < z^0 < z^1, \quad (z^1 - z^0) < \epsilon, \quad \langle z_n \rangle \rightsquigarrow (z^0, z^1).$$

It follows that

$$\langle z_n \cdot x_n \rangle \rightsquigarrow (z^0 \cdot x^0, z^1 \cdot x^1), \quad \langle z_n \cdot y_n \rangle \rightsquigarrow (z^0 \cdot y^0, z^1 \cdot y^1),$$

and the desired conclusion $Z \cdot X < Z \cdot Y$ will follow quite easily, if we could choose z^0, z^1 so that

$$z^1 x^1 < z^0 y^0,$$

or equivalently,

$$x^1(z^1 - z^0) < z^0(y^0 - x^1). \quad (\text{A.18})$$

Now (A.18) is obvious if $x^1 \leq 0$, because in that case $x^1(z^1 - z^0) \leq 0$ and $z^0(y^0 - x^1) > 0$. If $x^1 > 0$, we find first some $\delta > 0$ such that $\langle z_n \rangle \rightsquigarrow (\delta, \infty)$ and then z^0, z^1 such that

$$0 < \delta < z^0 < z^1, \quad \langle z_n \rangle \rightsquigarrow (z^0, z^1), \quad (z^1 - z^0) < \frac{\delta(y^0 - x^1)}{x^1}$$

which imply (A.18):

$$x^1(z^1 - z^0) < x^1 \frac{\delta(y^0 - x^1)}{x^1} \leq z^0(y^0 - x^1).$$

Verification of (A.16) is substantially simpler and completes the proof of the theorem. $\dashv$

A.49. Exercise. Prove that for all Dedekind cuts X, Y and Z :

$$X \cdot (Y + Z) = X \cdot Y + X \cdot Z.$$

(Use the formal definitions of $+$ and $\cdot$ given in the proof of **A.48**.)

A.50. Exercise. Prove (A.12) and (A.13).

A.51. Exercise. Show that if $\langle x_n \rangle$ is Cauchy in an ordered field and $\langle x_n \rangle \not\approx \langle 0 \rangle$, then the sequence $\langle y_n \rangle$ defined by (A.14) is also Cauchy and $\langle x_n y_n \rangle \approx \langle 1 \rangle$.

A.52. Theorem. Uniqueness of the real numbers. For any two complete, ordered fields F^1 and F^2 , there exists exactly one bijection

$$\pi^* : F^1 \rightarrow F^2$$

which is an isomorphism, i.e.

1. $\pi^*(0) = 0, \quad \pi^*(1) = 1,$
2. $\pi^*(x + y) = \pi^*(x) + \pi^*(y), \quad \pi^*(xy) = \pi^*(x)\pi^*(y),$
3. $x \leq y \iff \pi^*(x) \leq \pi^*(y).$

Proof. By the uniqueness of the rationals A.14, there exists (exactly one) isomorphism

$$\pi : Q^1 \rightarrow Q^2,$$

where Q^1, Q^2 are the sets of rationals in the two fields F^1, F^2 , and the problem is to extend this π to the whole of F^1 .

Lemma 1. For each $x \in F^1$, there exists a sequence $\langle x_n \rangle$ of rationals in F^1 , such that $\lim_n x_n = x$.

Proof. Using the density of the rationals (Exercise A.27), we can find for each $n \in N$ a rational $x_n \in Q^1$ such that $|x - x_n| < 1/(n+1)$, and then (easily, using problem A.26) $\lim_n x_n = x$.

Lemma 2. For each sequence $\langle x_n \rangle$ of rationals in F^1 ,

$$(\exists x \in F^1)[\lim_n x_n = x] \implies (\exists x^* \in F^2)[\lim_n \pi(x_n) = x^*].$$

Proof. We know that

$$u < v \iff \pi(u) < \pi(v) \quad (u, v \in Q^1)$$

because π is an isomorphism, so that for all $a, b \in Q^1$, $a < b$,

$$\langle x_n \rangle \rightsquigarrow (a, b) \iff \langle \pi(x_n) \rangle \rightsquigarrow (\pi(a), \pi(b)). \quad (\text{A.19})$$

If $\langle x_n \rangle$ converges, then it is Cauchy, so that $\langle \pi(x_n) \rangle$ is also Cauchy by (A.19) and A.36 (using A.26 once more), and therefore $\langle \pi(x_n) \rangle$ converges because F^2 is complete.

We can use the same simple idea to verify the third basic fact we need.

Lemma 3. For any two Cauchy sequences in the rationals of F^1 ,

$$[\lim_n x_n = \lim_n y_n] \implies \lim_n \pi(x_n) = \lim_n \pi(y_n).$$

The Lemmas guarantee that we can define unambiguously for each $x \in F^1$,

$$\pi^*(x) =_{\text{df}} \lim_n \pi(x_n), \quad \text{where } \lim_n x_n = x, \text{ with } x_n \in Q^1. \quad (\text{A.20})$$

Since for each rational $x \in Q^1$, $\lim_n x = x$, we have

$$\pi^*(x) = \lim_n \pi(x) = \pi(x),$$

so that π^* is an extension of π . It remains to verify that π^* is an isomorphism of F^1 with F^2 .

Suppose first that

$$x = \lim_n x_n, \quad y = \lim_n y_n, \quad x_n, y_n \in Q^1, \quad x < y.$$

This implies immediately (by **A.36**) that there exist rationals a, b, c and d satisfying

$$\langle x_n \rangle \rightsquigarrow (a, b), \quad \langle y_n \rangle \rightsquigarrow (c, d), \quad b < c,$$

and hence by (A.19)

$$\langle \pi(x_n) \rangle \rightsquigarrow (\pi(a), \pi(b)), \quad \langle \pi(y_n) \rangle \rightsquigarrow (\pi(c), \pi(d)).$$

It follows that

$$\pi^*(x) = \lim_n \pi(x_n) < \lim_n \pi(y_n) = \pi^*(y)$$

because $\pi(b) < \pi(c)$, and this completes the proof that π^* respects the relation $<$:

$$x < y \implies \pi^*(x) < \pi^*(y).$$

Directly from this, π^* is an injection and it respects the ordering,

$$x \leq y \iff \pi^*(x) \leq \pi^*(y).$$

The rest is trivial (if tiresome) and follows mostly from the *limit theorems* of the Calculus, which hold in every complete, ordered field—and can be proved easily with the tools we have developed. As an example:

$$\begin{aligned} \pi^*(x + y) &=_{\text{df}} \lim_n \pi(x_n + y_n), \quad \text{where } x_n \rightarrow x, y_n \rightarrow y, x_n, y_n \in Q^1, \\ &= \lim_n [\pi(x_n) + \pi(y_n)] \\ &= \lim_n \pi(x_n) + \lim_n \pi(y_n) \\ &= \pi^*(x) + \pi^*(y). \end{aligned}$$

The crucial step in this computation is the identity

$$\lim_n [\pi(x_n) + \pi(y_n)] = \lim_n \pi(x_n) + \lim_n \pi(y_n).$$

We skip the details. ⊥

A.53. Exercise. *Work out the details of the proof of*

$$\pi^*(x + y) = \pi^*(x) + \pi^*(y).$$

A.54. The real numbers. As we did for the natural numbers and the rationals, we now fix some complete, ordered field

$$(\mathcal{R}, 0, 1, +, \cdot, \leq), \tag{A.21}$$

whose members we will call **real numbers**. We emphasize once more that fixing some $\mathcal{R}$ is only a convenience and the specific choice of $\mathcal{R}$ is of no importance: the fundamental mathematical fact for the development of analysis is that a complete, ordered field exists and that any two complete, ordered fields are isomorphic.

A.55. Exercise. *Prove Corollary 2.14 in Chapter 2 from the axioms.*

The open sets of real numbers defined in **A.29** form a topology by the easy Exercise **A.30**, so we have notions of Borel sets of reals and Borel measurable functions from $\mathcal{R}$ to other topological spaces and vice versa, by **10.23** and **10.33**. The notion of Borel isomorphism was defined in **10.34**. The next theorem makes it possible to transfer results about Baire space to the reals, and it is the main tool for analyzing the set theoretic properties of $\mathcal{R}$. We will omit its proof, which is quite simple, using Problems ***x10.11** and **x10.12**.

A.56. Theorem. *As a topological space, $\mathcal{R}$ is Borel isomorphic with the Baire space $\mathcal{N}$.*

Problems

***xA.1.** Let F be the set of all rational functions with integer coefficients on $\mathcal{R}$, i.e. all real functions which can be represented as quotients of polynomials with integer coefficients and prove that it is a field with the obvious algebraic operations. Show that the relation

$$f \leq g \iff_{\text{df}} (\exists x)(\forall y \geq x)[f(y) \leq g(y)]$$

is an ordering on F , and with it F is a non-archimedean ordered field.

***xA.2.** Prove that there exists an ordered field which is not complete, but in which every Cauchy sequence has a limit. **HINT:** Show that every ordered field has a *Cauchy completion*.

xA.3. Every open set of reals is a countable union of disjoint open intervals.

xA.4. Every closed interval of real numbers $[a, b]$ is compact, in the topological sense, **10.37**.

xA.5. Every closed set of real numbers is a countable union of compact sets.

xA.6. Every closed set of real numbers F can be written uniquely as the disjoint union of a perfect and a countable set.

***xA.7.** Prove Theorem **A.56**.

Appendix B

AXIOMS AND UNIVERSES

The serious study of *models* of axiomatic set theories depends heavily on methods from *mathematical logic* which are outside the scope of these Notes.¹ Here we will consider only SET UNIVERSES, generalizations of the Zermelo and the Z-F universes of Chapter 11, which are very special models and can be studied by standard mathematical techniques, as we study fields or topological spaces. First we will prove that THE ZERMELO UNIVERSES of Chapter 11 ARE MODELS OF **ZDC** and THE Z-F UNIVERSES ARE MODELS OF **ZFDC**; this will give us a better understanding of these universes, and it will also yield some simple CONSISTENCY AND INDEPENDENCE RESULTS for the corresponding theories. In the main part of this chapter we will construct some new set universes with quite different properties, including the ANTIFOUNDED UNIVERSE of Aczel which contains a rich variety of ill founded sets. We will glean some consistency results from these models too, but consistency results are not our main concern: our primary interest is to explore and understand several natural, intuitive notions of SET and compare them with the standard conception of PURE, GROUNDED SET discussed in 12.25.

We begin with a result about the least Zermelo universe $\mathcal{Z}$ which is somewhat surprising, given how much we promoted $\mathcal{Z}$ in 11.25 as a rich collection of sets which contains all objects of interest of classical mathematics.

B.1. Theorem. *The set HF of all hereditarily finite sets is not a member*

¹For those who do know logic, we remark here that the most natural way to formalize the theories we have studied is in a *many sorted predicate logic with identity*, with separate variables for objects, definite conditions and definite operations of every arity, and relation symbols *Set* and $\in$. Notice that we did not assume in 3.18 any extensionality principles for conditions or operations, and we have never appealed to such principles. This means that to get (a notational variant of) the classical *Gödel-Bernays Theory* from the **ZFC** of 11.31 we must add *extensionality for conditions*. On the other hand, precisely because we did not include any extensionality axioms for conditions, we can view the present **ZFC** as a notational variant of the classical *Zermelo-Fraenkel Theory*, by interpreting “conditions” to be just “formulas with set parameters.”

of $\mathcal{Z}$; in fact, there is no set $A \in \mathcal{Z}$ such that

$$\emptyset \in A \ \& \ (\forall X)[X \in A \implies \mathcal{P}(X) \in A]. \quad (\text{B.1})$$

Proof. For each $x \in \mathcal{Z}$, we let

$$\text{level}(x) = \text{the least } n \text{ such that } x \in \mathcal{Z}_n, \quad (\text{B.2})$$

so that members of N_0 have level 0, but

$$\text{level}(x) > 0 \implies \text{level}(\{x\}) = \text{level}(x) + 1;$$

because $x \in \mathcal{Z}_n \implies \{x\} \in \mathcal{Z}_{n+1}$, and if $\{x\} \in \mathcal{Z}_n$ with $n > 0$, then $\{x\} \subseteq \mathcal{Z}_{n-1}$ by (11.20), which implies $x \in \mathcal{Z}_{n-1}$. Define now by recursion the sets

$$A_0 = \{\emptyset, \{\emptyset\}\}, \quad A_{n+1} = \{A_n\}.$$

Clearly, each $A_n \in HF$, $\text{level}(A_0) = 1$, since $A_0 \subseteq N_0$ but $A_0 \notin N_0$, and by induction, $\text{level}(A_n) = n + 1$.

Suppose now that there is some $A \in \mathcal{Z}$ which satisfies (B.1), and notice that $A_0 \in \mathcal{P}(\mathcal{P}(\emptyset)) \in A$, and then by induction, for each n , $A_n \in A$. If $A \in \mathcal{Z}_m$, then each $A_n \in \mathcal{Z}_m$ because $\mathcal{Z}_m$ is transitive, and hence $\text{level}(A_n) \leq m$ for each n , which violates what we just proved. $\dashv$

A similar argument shows that (with Kuratowski tupling) $\bigcup_{n=2}^{\infty} \{N_0\}^n \notin \mathcal{Z}$, Problem ***B.2**, and it is quite easy to extend it to show that $\mathcal{Z}$ misses many more very simple sets, but the fact that it lacks HF is undoubtedly the most startling of the lot. The construction of HF is so direct, it seems to follow so naturally from our most basic intuitions about sets, that it is really hard to believe that we developed all this set theory in Chapters 3 - 10 and Appendix A from axioms which do not guarantee its existence. One may try to write off this feature of the Zermelo axioms as a small oversight of Zermelo and strengthen the axioms in some minor way to ensure the existence of HF , but this is the wrong way to go. On the one hand, we know the natural extension of **ZAC**, it is the addition of the Replacement Axiom which can be justified by arguments only marginally different from those used to justify the Separation Axiom. And on the other hand, the importance of **ZAC** lies precisely in its two, contradictory features: that it can prove so much about classical mathematics (which is its real domain), while it can be interpreted in such simple, easy to comprehend models like $\mathcal{Z}$. Whatever doubts may have lingered about the soundness of set theory from the paradoxes should be at least moderated by the strength of **ZAC** and its ease of interpretation.

We still need to make precise the sense in which $\mathcal{Z}$ is a “model” of **ZDC**. Perhaps the simplest way to introduce the key, new idea we need, is to try and reinterpret **B.1** as an independence result.

B.2. Theorem? *We cannot prove in **ZDC** the proposition that some set A exists, which contains the empty set and is closed under the powerset operation.*

Proof. Spelled out symbolically for precision, the proposition in question is

$$\phi \iff_{\text{df}} (\exists A)[\emptyset \in A \ \& \ (\forall X)[X \in A \implies \mathcal{P}(X) \in A]].$$

Suppose, towards a contradiction, that ϕ is a theorem of **ZDC**. Since the least Zermelo universe $\mathcal{Z}$ has all the closure properties demanded by the axioms of **ZDC**, any proof of ϕ from these axioms could be translated into a proof of the interpretation of ϕ in $\mathcal{Z}$, which is

$$\phi^{(\mathcal{Z})} \iff (\exists A \in \mathcal{Z})[\emptyset \in A \ \& \ (\forall X \in \mathcal{Z})[X \in A \implies \mathcal{P}(X) \in A]].$$

As a consequence, $\phi^{(\mathcal{Z})}$ is true, so there exists a set $A \in \mathcal{Z}$ satisfying

$$\emptyset \in A \ \& \ (\forall X \in \mathcal{Z})[X \in A \implies \mathcal{P}(X) \in A],$$

which by the transitivity of $\mathcal{Z}$ is equivalent to

$$\emptyset \in A \ \& \ (\forall X)[X \in A \implies \mathcal{P}(X) \in A]; \quad (\text{B.3})$$

by **B.1**, no $A \in \mathcal{Z}$ satisfies (B.3). ⊥

The argument is unfamiliar, unless you know a lot of logic, and in any case it is incomplete, only a sketch. What we need to elucidate is the meaning of “interpreting a proposition in $\mathcal{Z}$,” the move from ϕ to $\phi^{(\mathcal{Z})}$ above, and it may help to consider an example. Suppose we have located a traditional mathematician (perhaps an old-fashioned analyst) who disclaims any interest in general set theory beyond its applications to classical mathematics, he has studied Chapters 3 - 10 and Appendix A and he is convinced that all the objects he cares about live in $\mathcal{Z}$. In an effort to simplify his world, he declares that henceforth by “object” he will mean “member of $\mathcal{Z}$,” that is his universe. Suppose further that this person now utters the Powerset Axiom and claims that he believes it. *What does he mean?* Spelled out in terms of the primitive notions of sethood and membership, the powerset axiom reads as follows:

(IV) : *For each object A , there exists a set B , such that for every X ,*

$$X \in B \iff \text{Set}(X) \ \& \ (\forall t)[t \in X \implies t \in A]. \quad (\text{B.4})$$

Replacing “object” by “member of $\mathcal{Z}$ ” in this, we get something rather different.

(IV)^($\mathcal{Z}$) : *For each $A \in \mathcal{Z}$, there exists some set $B \in \mathcal{Z}$, such that for every $X \in \mathcal{Z}$,*

$$X \in B \iff \text{Set}(X) \ \& \ (\forall t \in \mathcal{Z})[t \in X \implies t \in A]. \quad (\text{B.5})$$

This is what our friend really means when he tries to tell us that every set has a powerset, and it differs enough from the Powerset Axiom that its truth is not immediately apparent. To prove it, for each $A \in \mathcal{Z}$, we naturally take $B = \mathcal{P}(A)$, which is also in $\mathcal{Z}$ and satisfies (B.4), for every X . Notice next that for each $X \in \mathcal{Z}$,

$$(\forall t \in \mathcal{Z})[t \in X \implies t \in A] \iff (\forall t)[t \in X \implies t \in A],$$

easily, by the transitivity of $\mathcal{Z}$. This reduces (B.5) to

$$X \in B \iff \text{Set}(X) \ \& \ (\forall t)[t \in X \implies t \in A], \quad (\text{B.6})$$

which is true for every X and hence (in particular) for every $X \in \mathcal{Z}$.

As a matter of fact, all the axioms of **ZDC** yield true propositions when we replace in them “*object*” by “*member of $\mathcal{Z}$* ”. It follows that all the theorems derived from the axioms of **ZDC** *by logic alone* also yield true propositions when we understand them as assertions about $\mathcal{Z}$ in the same way, so our stipulated classical friend can safely work in **ZDC** and be assured that he is proving statements which are true of his world. This not entirely trivial proposition expresses the fact that the universe $\mathcal{Z}$ is a “model” of **ZDC**.

B.3. A **set universe** $\mathcal{M}$ is any triple M, S, E , of a class (which may be a set) M , a subclass $S \subseteq M$, and a binary definite condition E such that $E(x, y) \implies x, y \in M$ and for each $x \in M$,

$$\mathbf{b}_{\mathcal{M}}(x) =_{\text{df}} \{t \mid tEx\} \text{ is a set,} \quad (\text{B.7})$$

which means that for some set X and all t ,

$$t \in X \iff E(t, x).$$

We write synonymously

$$tEx \iff E(t, x)$$

for the condition E , which interprets the $\in$ condition in $\mathcal{M}$, and we call $\mathbf{b}_{\mathcal{M}}(x)$ the **body** of each $x \in M$. The $\mathcal{M}$ -**objects** are the members of M and the $\mathcal{M}$ -**sets** are the members of S . An n -ary definite condition is an $\mathcal{M}$ -**condition** if it only holds of objects in M , i.e.

$$P(x_1, \dots, x_n) \implies x_1, \dots, x_n \in M;$$

and a definite n -ary operation F is an $\mathcal{M}$ -**operation** if it assigns $\mathcal{M}$ -objects to $\mathcal{M}$ -objects, i.e.

$$x_1, \dots, x_n \in M \implies F(x_1, \dots, x_n) \in M.$$

A universe $\mathcal{M}$ is **natural** if the class M is transitive and S, E are the standard sethood and membership conditions, i.e.

$$\begin{aligned} X \in M &\implies X \subseteq M, \\ x \in S &\iff x \in M \ \& \ Set(x), \\ x E y &\iff x, y \in M \ \& \ x \in y. \end{aligned}$$

A natural universe $\mathcal{M}$ is completely determined by the transitive class M of its objects and we will identify it with that class, i.e. when we refer to *the universe M* for a transitive class M , we will mean the natural universe with objects the members of M . Notice that in a natural universe the body of each object is the set of its members,

$$\mathbf{b}_M(x) = \{t \mid t \in x\} \quad (x \in M); \quad (\text{B.8})$$

this means that $\mathbf{b}_M(x) = x$, if x is a set, but $\mathbf{b}_M(x) = \emptyset$ if x is an atom.

The **relativization** to a universe $\mathcal{M}$ of a proposition θ is the proposition $\theta^{(\mathcal{M})}$ constructed by replacing “object” by “ $\mathcal{M}$ -object”, “set” by “ $\mathcal{M}$ -set”, “ $x \in y$ ” by “ $x E y$ ”, “condition” by “ $\mathcal{M}$ -condition” and “operation” by “ $\mathcal{M}$ -operation”, consistently wherever these expressions occur in θ . If $\theta^{(\mathcal{M})}$ is true, we say that θ is **true in the universe $\mathcal{M}$** or (synonymously) that $\mathcal{M}$ **satisfies** or **models** θ .

A universe $\mathcal{M}$ is a **model** of an axiomatic system $\mathbf{T}$ if every axiom of $\mathbf{T}$ is true in $\mathcal{M}$, i.e. if the relativization $\theta^{(\mathcal{M})}$ of every axiom θ of $\mathbf{T}$ is a true proposition. If this holds, we call $\mathcal{M}$ a **universe of $\mathbf{T}$** , or simply a **$\mathbf{T}$ -universe**.

B.4. Propositions and relativizations. By *proposition* we mean any ordinary, definite mathematical statement or assertion, just like the axioms, hypotheses and theorems we have considered so far. This is not completely precise, everyday mathematical English not being a perfectly specified language. The basic idea is that for all objects x and y , the expressions $x = y$, $x \in y$ and $Set(x)$ are propositions; for each definite condition P and objects $x_1, \dots, x_n$, the expression $P(x_1, \dots, x_n)$ is a proposition; and that propositions may be combined by the basic operations of logic, $\neg$, $\vee$, $\exists$ and the like. Relativizations² to a set universe $\mathcal{M}$ are computed as one might

²To those who know formal logic, it might appear that the completely precise (syntactical) relativization operation on *formulas* is much easier to understand than this relativization operation, which is applied directly to propositions of “mathematical English.” But to apply the formal relativization process, we must first “formalize” the ordinary language propositions in which we are ultimately interested, and a moment’s thought will convince you that the formalization process is *exactly* as “vague” as the present operation of relativization. It may be argued that we know how to formalize a certain proposition precisely if we can interpret it in some arbitrary structure, i.e. precisely if we can understand its informal relativizations.

expect, e.g.

$$\begin{aligned}
 (x = y)^{(\mathcal{M})} &: x = y, \\
 \text{Set}(x)^{(\mathcal{M})} &: x \in S, \\
 (x \in y)^{(\mathcal{M})} &: x E y, \\
 P(x_1, \dots, x_n)^{(\mathcal{M})} &: P(x_1, \dots, x_n), \\
 (\neg \theta)^{(\mathcal{M})} &: \neg(\theta^{(\mathcal{M})}), \\
 (\phi \ \& \ \psi)^{(\mathcal{M})} &: \phi^{(\mathcal{M})} \ \& \ \psi^{(\mathcal{M})}, \\
 ((\forall x)\phi)^{(\mathcal{M})} &: (\forall x \in M)\phi^{(\mathcal{M})}, \\
 ((\forall P)\phi)^{(\mathcal{M})} &: (\forall P)[(\forall x, y)[P(x, y) \implies x, y \in M] \implies \phi^{(\mathcal{M})}],
 \end{aligned}$$

where P varies over binary, definite conditions, etc. In the proofs which follow, we will take care to spell out laboriously the relativizations of all the propositions that concern us. This will produce many examples which illustrate the notion, and it will also ensure that the specific results we claim will be precise and rigorously established even if the general notion of proposition and the relativization process are not precisely delimited. On the other hand, much of the significance of the results comes from the following general principle. It says simply that logical consequences of true (in $\mathcal{M}$) propositions are true (in $\mathcal{M}$) for any universe $\mathcal{M}$, and, of course, it holds for arbitrary models of arbitrary axiomatic theories, not just set universes.

B.5. Principle of Soundness of Logical Inference. *If a proposition θ is a theorem of an axiomatic system $\mathbf{T}$ (i.e. it can be proved by logic alone from the axioms of $\mathbf{T}$), then every universe of $\mathbf{T}$ satisfies θ .*

B.6. Universes vs. general models. According to the discussion in 8.20, to define a model of an axiomatic set theory we must specify a domain of objects, define on it the conditions of membership and sethood and also specify which conditions and operations on the domain will be considered definite. Set universes are very special models in two ways.

(1) When we view a set universe $\mathcal{M}$ as a model for an axiomatic set theory, we take its *definite conditions* to be all the definite conditions of our basic domain $\mathcal{W}$, and we take for its *definite operations* all the $\mathcal{M}$ -operations, i.e. the definite operations of $\mathcal{W}$ which take $\mathcal{M}$ -objects to $\mathcal{M}$ -objects.³ It is routine to verify that all the axioms for definite conditions and operations listed in 3.18 hold with this interpretation, Problem xB.1:

³More pedantically, the definite conditions of $\mathcal{M}$ are the restrictions to M of the definite conditions in the intended universe $\mathcal{W}$, and similarly for the operations.

thus, to prove that a set universe $\mathcal{M}$ is a model of (say) **ZFDC**, it is enough to prove the relativizations to $\mathcal{M}$ of axioms **(I)** - **(VIII)**.

(2) Because we assume that the body $\mathbf{b}_{\mathcal{M}}(x)$ of each $x \in M$ is a set, $\mathcal{M}$ -sets cannot be “larger” than the sets of the intended universe $\mathcal{W}$, for example, there cannot be an $\mathcal{M}$ -set x such that for all t , tEx .

Natural universes are even more special, of course, they only restrict the domain of objects to some transitive class—and this makes it especially simple to understand the meaning of the relativization operation for them. From the mathematical point of view, natural universes are *subuniverses* of $\mathcal{W}$ and they inherit their structure from $\mathcal{W}$, much like subgroups, subposets, topological subspaces and the like are specified by a subset of some given space and inherit the relevant structure from it. The additional subtlety here is that we need to interpret (relativize) in these subuniverses propositions which are logically quite complex, more complex than the typical identities or inclusions which come up in Algebra or Topology.

We begin with the verification of the axioms we have been studying in natural universes which we have already introduced, where the notions are most familiar.

B.7. Lemma. *Every transitive class M satisfies the Axiom of Extensionality.*

Proof. The relativization to M of the Extensionality Axiom reads as follows:

(I)^(M) : For all sets $A, B \in M$,

$$A = B \iff (\forall x \in M)[x \in A \iff x \in B].$$

If $A = B$, then A and B have the same members, so they certainly have the same members in M . To prove the converse, we must show that if $A \neq B$, then there must exist some $x \in M$, such that either $x \in A \setminus B$ or $x \in B \setminus A$; but if $A \neq B$, then there certainly exists some

$$x \in (A \setminus B) \cup (B \setminus A) \subseteq A \cup B,$$

$A, B \subseteq M$ because M is transitive, and hence this x is also in M . ⊣

B.8. Lemma. *Every Zermelo universe M is a natural universe of axioms **(I)** - **(VI)**.*

Proof. We consider in turn each of the axioms **(I)** - **(VI)** other than **(I)** just shown and **(IV)** for which we gave the argument in the example above. The reader is advised to compare the relativization of each axiom to M

which we must prove with the original statement of the axiom in Chapter 3.

(II)^(M) Emptyset and Pairset: (a) *There is a special object $y \in M$ which is a set but has no members in M .* (b) *For all $x, y \in M$, there is a set $z \in M$ such that*

$$(\forall t \in M)[t \in z \iff t = x \vee t = y]. \quad (\text{B.9})$$

(a) $\emptyset \in M$ by hypothesis, it has no members whatsoever, so it certainly has no members in M . (b) If $x, y \in M$, then $z = \{x, y\} \in M$ by hypothesis, and it obviously satisfies (B.9), since it satisfies the stronger

$$(\forall t)[t \in z \iff [t = x \vee t = y]].$$

(III)^(M) Separation Axiom: *For each $A \in M$ and each unary, definite condition P , there exists a set $B \in M$ which satisfies the equivalence*

$$(\forall x \in M)[x \in B \iff x \in A \ \& \ P(x)]. \quad (\text{B.10})$$

Suppose $A \in M$. By the Axiom of Separation, there exists some B which satisfies

$$(\forall x)[x \in B \iff x \in A \ \& \ P(x)]. \quad (\text{B.11})$$

Now $B \subseteq A$, so $B \in \mathcal{P}(A) \in M$, so $B \in M$ because M is transitive, and (B.11) implies the weaker (B.10).

(V)^(M) Unionset Axiom: *For each $\mathcal{E} \in M$, there exists a set $B \in M$ which satisfies the equivalence*

$$(\forall t \in M)[t \in B \iff (\exists X \in M)[X \in \mathcal{E} \ \& \ t \in X]]. \quad (\text{B.12})$$

Again, we naturally take $B = \bigcup \mathcal{E}$, which is in M by hypothesis and satisfies the equivalence

$$(\forall t)[t \in B \iff (\exists X)[X \in \mathcal{E} \ \& \ t \in X]]. \quad (\text{B.13})$$

Using once more the transitivity of M , immediately, for every t ,

$$(\exists X)[X \in \mathcal{E} \ \& \ t \in X] \iff (\exists X \in M)[X \in \mathcal{E} \ \& \ t \in X],$$

so that (B.12) reduces to

$$(\forall t \in M)[t \in B \iff (\exists X)[X \in \mathcal{E} \ \& \ t \in X]],$$

which is implied by the stronger (B.13).

(VI)^(M) Axiom of Infinity: *There exists a set $I \in M$ such that*

$$\emptyset \in I \ \& \ (\forall x \in M)[x \in I \implies \{x\} \in I].$$

This is quite simple, taking $I = N_0 \in M$.

⊣

B.9. Theorem. (1) *Every Zermelo universe is a natural universe of **ZDC**, and every Z-F universe is a natural universe of **ZFDC**.*

(2) **(AC)** *Every Zermelo universe is a natural universe of **ZAC**, and every Z-F universe is a natural universe of **ZFAC**.*

*In particular, $\mathcal{Z}$ and every $M(I)$ such that $N_0 \subseteq I$ are natural universes of **ZDC**, or universes of **ZAC**, granting **AC**.*

(3) (von Neumann) **(AC)** *The von Neumann universe $\mathcal{V}$ is a natural universe of **ZFC**.*

Proof. The relativizations of **DC** and **AC** were proved in 11.23 and the relativization of the Axiom of Replacement is exactly the defining condition of a Z-F universe in 11.33. For Part (3), we have already shown that $\mathcal{V}$ is a Z-F universe in 11.34, so it only remains to prove the relativizations to $\mathcal{V}$ of the Principles of Purity 3.24 and Foundation 11.29. The first of these is

$\text{Purity}^{(\mathcal{V})}$: For every $x \in \mathcal{V}$, $\text{Set}(x)$,

and it is true simply because every object in $\mathcal{V}$ is a set, and the interpretation of “sethood” in $\mathcal{V}$ is the standard one. For the second, it is easiest to relativize the elementary version of the Foundation Principle in 11.30.

$\text{Foundation}^{(\mathcal{V})}$: For every set $X \in \mathcal{V}$, there exists some $m \in X$ such that $m \cap X$ is empty in $\mathcal{V}$, i.e.

$$(\forall t \in \mathcal{V})[t \notin m \vee t \notin X].$$

The negation of this would give us an $X \in \mathcal{V}$ and some $a \in X$ such that

$$(\forall m \in X)(\exists t \in X)[t \in m],$$

which by **DC** (as in the proof of 11.30) implies that X is ill founded, contradicting the assumption $X \in \mathcal{V}$. $\dashv$

The Soundness of Logical Inference **B.5** combines well with **B.9** to yield many simple but interesting independence results about **ZDC** and **ZAC**: to prove that a proposition θ cannot be a theorem of **ZDC**, it is enough to find some $I \supseteq N_0$ such that $\theta^{(M(I))}$ is false. This is exactly the way we proved **B.2**, a bit clumsily without the precise notions. We have included in the problems several examples of this kind.

By the same reasoning, Part (3) of Theorem **B.9** implies that *we cannot refute in **ZFAC** the Principles of Purity or Foundation*, because $\mathcal{V}$ is a model of **ZFAC** which satisfies these principles. It should also be obvious that we cannot prove these principles in **ZFAC**, but to establish this rigorously we need to construct universes of **ZFAC** which are not natural. The basic tool for such constructions is the next simple notion.

B.10. Definition. A **Rieger universe** is any set universe $\mathcal{M} = M, S, E$ such that for every set $Y \subseteq M$, there exists exactly one $\mathcal{M}$ -set X satisfying $Y = \mathbf{b}_{\mathcal{M}}(X)$. For each $Y \subseteq M$ we set

$$\rho_{\mathcal{M}}(Y) =_{\text{df}} \text{the unique } X \in S \text{ such that } \mathbf{b}_{\mathcal{M}}(X) = Y, \quad (\text{B.14})$$

so that $\rho_{\mathcal{M}}$ is a definite operation and immediately from its definition, for every $Y \subseteq M$,

$$X = \rho_{\mathcal{M}}(Y) \iff X \in S \ \& \ \mathbf{b}_{\mathcal{M}}(X) = Y, \quad (\text{B.15})$$

$$t E \rho_{\mathcal{M}}(Y) \iff t \in Y. \quad (\text{B.16})$$

B.11. Rieger's Theorem. Every Rieger universe is a universe of **ZFDC**.

Proof. Fix $\mathcal{M} = M, S, E$ with the Rieger property and let $\mathbf{b}(x) = \mathbf{b}_{\mathcal{M}}(x)$, skipping the subscript since $\mathcal{M}$ is the only universe around. We verify in turn the relativizations of all the axioms of **ZFDC**.

(I)^(M) Extensionality Axiom: For all $A, B \in M$, if $S(A)$ and $S(B)$, then

$$A = B \iff (\forall x \in M)[x E A \iff x E B].$$

If $A = B$, then surely, for all x , $x E A \iff x E B$. Conversely, if for all $x \in M$, $x E A \iff x E B$, then $\mathbf{b}(A) = \mathbf{b}(B)$; by the Rieger property, there is exactly one $C \in S$ such that $\mathbf{b}(A) = \mathbf{b}(C)$, so we must have $C = A$, and similarly $C = B$, hence $A = B$.

(II)^(M) Emptyset and Pairset: (a) There is a special object $y \in S$ such that $(\forall t \in M) \neg t E y$. (b) For all $x, y \in M$, there is some $z \in S$ such that

$$(\forall t \in M)[t E z \iff t = x \vee t = y].$$

For (a), choose y so that $y \in S$ and $\mathbf{b}(y) = \emptyset$, and for (b) choose $z \in S$ so that $\mathbf{b}(z) = \{x, y\}$, both times by applying directly the Rieger property. In the case of (b), for example, we compute:

$$t E z \iff t \in \mathbf{b}(z) \iff [t = x \vee t = y],$$

which is the required conclusion.

(IV)^(M) Powerset Axiom: For each $A \in M$, there exists some $B \in S$, such that for every $X \in M$,

$$X E B \iff X \in S \ \& \ (\forall t \in M)[t E X \implies t E A].$$

Given $A \in M$, choose $B \in S$ by the Rieger property so that

$$\mathbf{b}(B) = \{\rho(Y) \mid Y \subseteq \mathbf{b}(A)\}, \quad (\text{B.17})$$

where $\rho(Y) = \rho_{\mathcal{M}}(Y)$ is the Rieger operation associated with $\mathcal{M}$ by (B.14), and compute:

$$\begin{aligned}
 XEB &\iff X \in \mathbf{b}(B) \\
 &\iff (\exists Y)[Y \subseteq \mathbf{b}(A) \ \& \ X = \rho(Y)] \\
 &\iff (\exists Y)[Y \subseteq \mathbf{b}(A) \ \& \ [X \in S \ \& \ \mathbf{b}(X) = Y]] \quad \text{by (B.15)} \\
 &\iff X \in S \ \& \ \mathbf{b}(X) \subseteq \mathbf{b}(A) \\
 &\iff X \in S \ \& \ (\forall t \in M)[tEX \implies tEA].
 \end{aligned}$$

Verifications of the remaining axioms in $\mathcal{M}$ are similar, following the same ideas as in the proof of Theorem B.9, and we leave them for the exercises. $\dashv$

B.12. Exercise. *Prove that a set universe $\mathcal{M} = M, S, E$ is a Rieger universe if and only if (1) it satisfies the Axiom of Extensionality, and (2) for every $Y \subseteq M$, there exists some $X \in S$ such that $\mathbf{b}_{\mathcal{M}}(X) = Y$.*

B.13. Exercise. *Every Rieger universe is a model of the Axioms of Separation (III) and Replacement (VIII).*

B.14. Exercise. (AC) *Every Rieger universe is a model of ZFAC.*

B.15. Relativization of “faithfully modeled” notions. The Choice Principles DC and AC are formulated in terms of the notions of “function” which was defined in Chapter 4 using an “arbitrary but fixed” ordered pair operation 4.4, and “system of natural numbers,” which was also “arbitrary but fixed” in 5.9. It is not completely obvious how to relativize propositions involving such “faithfully modeled” notions, since (for example) a given universe $\mathcal{M}$ may not be closed under the chosen ordered pair operation. We avoided the problem in 11.23 and B.9 by assuming that the fixed ordered pair is the Kuratowski pair under which every universe satisfying (I) - (VI) is closed, but there may be some lingering vagueness on how to deal with this problem in general. There is an easy solution for Rieger universes, which we outline in Problems **xB.13** and ***xB.14**. In discussing Rieger universes from now on, we will assume tacitly that we have fixed an ordered pair operation, a system of natural numbers, etc. for each of them, and we will relativize propositions which involve these faithfully modeled notions in terms of these fixed operations. Problems **xB.13** and ***xB.14** make it clear that which particular definitions are chosen is irrelevant for the results.

B.16. Proposition. *There exists a Rieger universe $\mathcal{M}_a$ in which every set is equinumerous with a set of atoms; in particular, we cannot prove in ZFDC that every set is pure.*

Proof. The idea is to code every set A by the pair $(0, A)$ in $\mathcal{M}_a$, and to declare that every object which does not code a set in this way is an atom. We set

$$\begin{aligned} x \in M_a &\iff_{\text{df}} x = x, \text{ so } M_a = \mathcal{W}, \\ x \in S_a &\iff_{\text{df}} (\exists A)[\text{Set}(A) \ \& \ x = (0, A)], \\ x E_a(0, A) &\iff_{\text{df}} x \in A, \end{aligned}$$

and proceed to verify that $\mathcal{M}_a = M_a, S_a, E_a$ is a set universe, that it has the Rieger property and that it satisfies the proposition “every set is equinumerous with a set of atoms.”

Skipping the subscript $\mathcal{M}_a$, clearly

$$\mathbf{b}(x) = \begin{cases} \emptyset, & \text{if for all sets } A, x \neq (0, A), \\ A, & \text{if for some (necessarily unique) set } A, x = (0, A), \end{cases}$$

so each $\mathbf{b}(x)$ is a set. For the Rieger property, we notice first that for each set $A \subseteq M_a$, $(0, A) \in S_a$ and $\mathbf{b}((0, A)) = A$; if $x \in S$ and $\mathbf{b}(x) = A$, then $x = (0, B)$ and $\mathbf{b}(x) = B$ for some set B , by the definition, so we have $A = B$ and $x = (0, A)$, as required. The Rieger operation is very simple in this case,

$$\rho(Y) = (0, Y).$$

The relativization to $\mathcal{M}_a$ of the proposition “every set is equinumerous with a set of atoms” is the assertion that *for every $\mathcal{M}_a$ -set $(0, A)$, there is a bijection in $\mathcal{M}_a$ between $(0, A)$ and some $\mathcal{M}_a$ -set of atoms $(0, B)$* . We take

$$B =_{\text{df}} \{(1, t) \mid t \in A\};$$

now $(0, B) \in S_a$ and, by the definition,

$$x E_a(0, B) \implies x \in B \implies (\forall y)[x \neq (0, y)] \implies x \notin S_a,$$

i.e. each $\mathcal{M}_a$ -member of $(0, B)$ is an atom in $\mathcal{M}_a$. If $(x, y)_a$ is the ordered pair operation of $\mathcal{M}_a$ and

$$f =_{\text{df}} \rho\{(t, (1, t))_a \mid t \in A\},$$

then $\mathcal{M}_a$ recognizes f as a bijection between the $\mathcal{M}_a$ -sets $(0, A)$ and $(0, B)$. $\dashv$

By the anthropomorphic “ $\mathcal{M}_a$ recognizes $f \dots$ ” we simply mean that if $A' = (0, A)$ and $B' = (0, B)$ are the objects in $\mathcal{M}_a$ which code the sets A and B , respectively, and if we set

$$\begin{aligned} \theta \iff_{\text{df}} f \subseteq A \times B \ \& \ (\forall x \in A')(\exists y \in B')(x, y) \in f \\ \ \& \ (\forall y \in B')(\exists! x \in A')[(x, y) \in f], \end{aligned}$$

then the relativization $\theta^{\mathcal{M}_a}$ of the proposition θ to $\mathcal{M}_a$ is true. This relativization can be computed in principle, but it is quite messy. It is best to develop a machinery for arguing about relativizations without actually writing them out, and for this the following, traditional “model theoretic” notation is very useful. For each set universe $\mathcal{M}$ and each proposition θ ,

$$\mathcal{M} \models \theta \iff_{\text{df}} \theta \text{ is true in } \mathcal{M} \iff \theta^{\mathcal{M}}. \quad (\text{B.18})$$

We read $\mathcal{M} \models \theta$ “ $\mathcal{M}$ models θ ”, but also “ $\mathcal{M}$ thinks that θ ”, “ $\mathcal{M}$ believes that θ ”, etc. as befits the occasion. For example, for each pair of classes M , S and each binary condition E , let

$$\begin{aligned} \text{Setuniv}(M, S, E) \iff_{\text{df}} & (\forall u, v)[u E v \implies u, v \in M] \\ & \& (\forall t)[t \in S \implies t \in M] \\ & \& (\forall x)[x \in M \implies \\ & \quad (\exists X)[\text{Set}(X) \& (\forall t)[t \in X \iff t E x]]] \end{aligned}$$

be the fairly complex proposition which asserts that M , S , E comprise a set universe. Consider also

$$\begin{aligned} \text{Rieger}(M, S, E) \iff_{\text{df}} & (\forall Y)[(\text{Set}(Y) \& Y \subseteq M) \\ & \implies (\exists! X \in S)[Y = \mathbf{b}_{\mathcal{M}}(X)]] \\ \iff & (\forall Y)[(\text{Set}(Y) \& Y \subseteq M) \\ & \implies (\exists! X \in S)(\forall t \in M)[t \in Y \iff t E X]], \end{aligned}$$

which asserts that $\mathcal{M}$ is a Rieger universe. These are propositions about M , S and E , which may be true or false; whether they are true or not, it makes sense to interpret them (relativize them) in some universe $\mathcal{M}'$ and ask if $\mathcal{M}'$ *thinks that* M , S , E comprise a Rieger universe! There is an obvious question here, which has a simple and useful answer.

B.17. Theorem. *Suppose $\mathcal{M}_1 = M_1, S_1, E$ is a Rieger universe, $S_2 \subseteq M_2 \subseteq M_1$ are classes and E_2 is a binary condition such that $x E_2 y \implies x, y \in M_2$; if*

$$\mathcal{M}_1 \models \text{Setuniv}(M_2, S_2, E_2) \& \text{Rieger}(M_2, S_2, E_2),$$

then $\mathcal{M}_2 = M_2, E_2, E_2$ is also a Rieger universe.

Proof. To show first that $\mathcal{M}_2$ is a set universe, we must verify that for every $x \in M_2$,

$$(\exists Y \in \text{Set})(\forall t)[t \in Y \iff t \in M_2 \& t E_2 x]. \quad (\text{B.19})$$

Fix some $x \in M_2$. The proposition (B.19) is true in $\mathcal{M}_1$ since

$$\mathcal{M}_1 \models \text{Setuniv}(M_2, S_2, E_2),$$

which means that some $Y_1 \in S_1$ exists such that for all $t \in M_1$,

$$t E_1 Y_1 \iff t \in M_2 \ \& \ t E_2 x; \quad (\text{B.20})$$

and since $\mathcal{M}_1$ is a set universe, there exists some set Y such that for all t ,

$$t \in Y \iff t \in M_1 \ \& \ t E_1 Y_1; \quad (\text{B.21})$$

now (B.20) and (B.21) together imply what (B.19) demands for the given $x \in M_2$ and some Y .

To show that $\mathcal{M}_2$ is a Rieger universe, we must show first that if $Y \subseteq M_2$, then there exists some $X_2 \in S_2$ such that for every $t \in M_2$,

$$t \in Y \iff t E_2 X_2, \quad (\text{B.22})$$

and then verify that this X_2 is unique. Since $Y \subseteq M_2 \subseteq M_1$ and $\mathcal{M}_1$ is a Rieger universe, there exists some $X_1 \in S_1$ such that for all $t \in M_1$,

$$t \in Y \iff t E_1 X_1; \quad (\text{B.23})$$

working in $\mathcal{M}_1$, we apply the Rieger property for $\mathcal{M}_2$ to the $\mathcal{M}_1$ -set X_1 , to get some $X_2 \in S_2$ such that

$$\mathcal{M}_1 \models (\forall t)[t \in M_2 \implies [t \in X_1 \iff t E_2 X_2]]; \quad (\text{B.24})$$

and computing the relativization, this means that

$$(\forall t \in M_1)[t \in M_2 \implies [t E_1 X_1 \iff t E_2 X_2]]. \quad (\text{B.25})$$

Compute now, for any $t \in M_2 \subseteq M_1$:

$$\begin{aligned} t \in Y &\iff t E_1 X_1 \quad \text{by (B.23),} \\ &\iff t E_2 X_2 \quad \text{by (B.25),} \end{aligned}$$

which proves (B.22). The fact that at most one $X_2 \in S_2$ can satisfy (B.22) for every $t \in M_2$ is proved similarly. $\dashv$

REMARK: Notice that in relativizing (B.19) in this proof, we left the clause $t \in M_2$ alone. In computing relativizations, “primitive” propositions of the form $P(x_1, \dots, x_n)$ (which express that a definite condition P holds of the objects $x_1, \dots, x_n$) are their own relativizations.

This simple theorem makes it possible to construct universes within universes within universes, each time using the properties of the model just constructed. Consider, for example, the next Corollary of **B.16**.

B.18. Proposition. *There exists a Z-F universe $\mathcal{M}$ which has exactly one atom.*

Proof. Suppose c is an atom and let

$$M_c =_{\text{df}} \{x \mid (\forall t \in TC(x))[\neg \text{Set}(t) \implies t = c]\} \quad (\text{B.26})$$

be the class of objects *supported by* $\{c\}$ in the sense of Problem **x11.21**. It is quite easy to verify that M_c is transitive and for each set X ,

$$X \subseteq M_c \implies X \in M_c,$$

so the natural universe M_c has the Rieger property and by **B.11** it is a model of **ZFDC**; and it is quite clear that it has exactly one atom, c .

So far so good, as long as there exists at least one atom, which may or may not be true in our intended domain $\mathcal{W}$. But there are lots of atoms in the Rieger universe $\mathcal{M}_a$ of **B.16**, so what we need to do is *to interpret the proof of the preceding paragraph in the universe $\mathcal{M}_a$* . This argument runs as follows.

Let

$$\begin{aligned} \phi(M, S, E, c) \iff_{\text{df}} & \text{Setuniv}(M, S, E) \ \& \ \text{Rieger}(M, S, E) \quad (\text{B.27}) \\ & \& \neg \text{Set}(c) \ \& \ (\forall t \in M)[t \notin S \implies t = c] \end{aligned}$$

be the proposition which asserts of M, S, E, c that they have the properties we are interested in, and let

$$\theta \iff_{\text{df}} (\exists M)(\exists S)(\exists E)(\exists c)\phi(M, S, E, c) \quad (\text{B.28})$$

be the proposition which asserts that some M, S, E, c with these properties exist. We have proved θ from the hypothesis that some atom exists, and other than that we have only used the axioms of **ZFDC**—what else is there! Hence this θ is true in every universe of **ZFDC** which has an atom, in particular $\mathcal{M}_a$, i.e.

$$\mathcal{M}_a \models \theta.$$

This means that for some $\mathcal{M}_a$ -classes M, S , some binary $\mathcal{M}_a$ -condition E and some $\mathcal{M}_a$ -object c ,

$$\mathcal{M}_a \models \phi(M, S, E, c),$$

and in particular

$$\mathcal{M}_a \models \text{Setuniv}(M, S, E) \ \& \ \text{Rieger}(M, S, E),$$

so by **B.17**, $\mathcal{M} = M, S, E$ is a Rieger universe. In addition

$$\mathcal{M}_a \models \neg \text{Set}(c) \ \& \ (\forall t \in M)[t \notin S \implies t = c],$$

which means precisely that

$$\mathcal{M} \models \neg \text{Set}(c) \ \& \ (\forall t)[\neg \text{Set}(t) \implies t = c],$$

the required conclusion that $\mathcal{M}$ believes that exactly one atom exists. $\dashv$

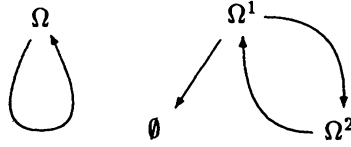


Figure B.1. Two decorated, ill-founded graphs.

It is not hard to manufacture Rieger universes with various types of ill founded sets, by a combination of the techniques in **B.16** and **B.17**. Some of the problems are about such results. Here we will concentrate on the construction of Aczel's *Antifounded Universe* $\mathcal{A}$, which has a rich variety of ill founded sets with well understood structure.

The idea for $\mathcal{A}$ comes from the Mostowski Collapsing Lemma **11.36**, which gives a “structural” characterization of pure, grounded sets. Recall that by **11.35**, a decoration of a graph G is any surjection $d : G \rightarrow d[G]$ such that

$$d(x) = \{d(y) \mid y \leftarrow x\} \quad (x \in G), \quad (\text{B.29})$$

where $\rightarrow$ is the edge relation on G and $\leftarrow$ is its inverse,

$$y \leftarrow x \iff y \text{ is a child of } x \iff x \rightarrow y.$$

Each grounded graph G admits a unique decoration d_G , and the pure, grounded sets are all the values $d_G(x)$ of these decorations. Can we also “decorate” the nodes of ill founded graphs to get pure, ill founded sets which are related to ill founded graphs in the same way that pure, grounded sets are related to grounded graphs?

B.19. Antifoundation Principle, AFA. *Every graph admits a unique decoration.*

In Figure B.1 we have labeled the nodes of two ill founded graphs by the values of their unique decorations, assuming that such exist. By the definition of decoration,

$$\Omega = \{\Omega\}, \quad \Omega^1 = \{\emptyset, \Omega^2\}, \quad \Omega^2 = \{\Omega^1\}, \quad (\text{B.30})$$

i.e. Ω , Ω^1 and Ω^2 are the “ultimately frustrating gifts” we discussed in **11.32**. We can refer to “the” frustrating gifts, because, in fact, the equations in (B.30)—or the graphs in Figure B.1—characterize these sets under **AFA**, as follows.

B.20. Proposition. (AFA) (1) *There is exactly one set Ω which is its own singleton.* (2) *There is exactly one pair of sets Ω^1, Ω^2 such that $\Omega^1 = \{\emptyset, \Omega^2\}$ and $\Omega^2 = \{\Omega^1\}$.*

Proof. (1) If $X = \{X\}$ and $Y = \{Y\}$, then we can use either X or Y to decorate the single node graph in Figure B.1; but this graph has only one decoration by **AFA**, so $X = Y$. The proof of (2) is similar. $\dashv$

This “uniqueness” part of the Antifoundation Principle we just applied makes it possible to specify and analyze the structure of ill founded sets with diverse properties, and is the main advantage of the antifounded universe $\mathcal{A}$ over other models which contain ill founded sets. We now proceed to its construction.

B.21. Definition. A **pointed graph** is a pair (G, p_G) of a graph and a node in it, in full detail, a structured set $(G, \rightarrow_G, p_G)$ where $p_G \in G$ and $\rightarrow_G$ is a binary relation on the field G . The designated node p_G is the **point** of the pointed graph.

B.22. Pictures. A pointed graph (G, p) is a **picture** of a set A , if there exists a decoration $d : G \rightarrow d[G]$ of G such that $d_G(p) = A$. The **canonical picture** of a pure set A is the pointed graph $(TC(A), \ni, A)$, where $TC(A)$ is the transitive closure of A and $\ni$ is the restriction of the inverse membership condition to $TC(A)$. This is a picture of A , because the identity function $d(x) = x$ is obviously a decoration of it, $A \in TC(A)$ and $d(A) = A$.

B.23. An **isomorphism** between two graphs G and H is any bijection $\pi : G \rightarrow H$ such that

$$x \rightarrow_G y \iff \pi(x) \rightarrow_H \pi(y) \quad (x, y \in G),$$

and an isomorphism between two pointed graphs (G, p) and (H, q) is a graph isomorphism $\pi : G \rightarrow H$ such that $\pi(p) = q$. We call G **isomorphic** with H if there exists an isomorphism $\pi : G \rightarrow H$ of the appropriate kind.

It is easy to construct non-isomorphic pointed graphs which picture the same set, even grounded ones, e.g. see Figure B.2 where we have labelled the nodes with the values of the unique decorations. On the other hand, we would expect that if a pointed graph (G, p) admits a unique decoration d_G , then the set $A = d_G[G]$ captures some important invariant of (G, p) . The next fundamental definition captures that invariant.

B.24. Definition. A relation $R \subseteq G \times H$ is a **bisimulation** between two pointed graphs $(G, \rightarrow_G, p_G)$ and $(H, \rightarrow_H, p_H)$, if it relates the points, i.e. $p_G R p_H$ and also satisfies the implication

$$\begin{aligned} x R y \implies & (\forall u \leftarrow_G x)(\exists v \leftarrow_H y)u R v \\ & \& (\forall v \leftarrow_H y)(\exists u \leftarrow_G x)u R v. \end{aligned} \tag{B.31}$$

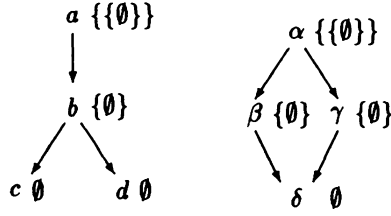


Figure B.2. Non-isomorphic, bisimilar, grounded graphs.

Two pointed graphs G, H are **bisimilar** if some bisimulation between them exists,

$$G =_{\text{bs}} H \iff_{\text{df}} (\exists R \subseteq G \times H)[R \text{ is a bisimulation}]. \quad (\text{B.32})$$

As usual with structured sets, we will often refer to “a pointed graph G ,” skipping the explicit reference to the edge relation $\rightarrow_G$ or the point when it is obvious or irrelevant—we already did this in (B.32).

B.25. Exercise. *Isomorphic pointed graphs are bisimilar, and so are the non-isomorphic, grounded, pointed graphs G and H in Figure B.2.*

B.26. Exercise. *If a is a minimal node in a graph G and b is a minimal node in H , then $\{(a, b)\}$ is a bisimulation of the pointed graphs (G, a) and (H, b) .*

B.27. Exercise. *Let L be the “single loop” graph on a singleton $\{a\}$, with the one edge pair (a, a) , and on the set of integers N define the successor edge relation*

$$n \rightarrow_s m \iff_{\text{df}} n + 1 = m.$$

Show that the relation $\{(a, i) \mid i \in N\}$ is a bisimulation of (L, a) with (N, n) , for every n .

B.28. Lemma. *The condition $=_{\text{bs}}$ is an equivalence condition on the class of all pointed graphs.*

Proof. Of the three properties of an equivalence condition (defined in 12.30), only the transitivity of $=_{\text{bs}}$ is not immediate. To prove that, suppose G_1, G_2 and G_3 are pointed graphs, R_1 is a bisimulation of G_1 with G_2 and R_2 is a bisimulation of G_2 with G_3 , and let

$$x R z \iff_{\text{df}} (\exists y)[x R_1 y \ \& \ y R_2 z] \quad (\text{B.33})$$

be the “product relation” of R_1 and R_2 . It is clear that R relates the points p_1 and p_3 of G_1 and G_3 , because $p_1 R_1 p_2$ and $p_2 R_2 p_3$ hold. Suppose that

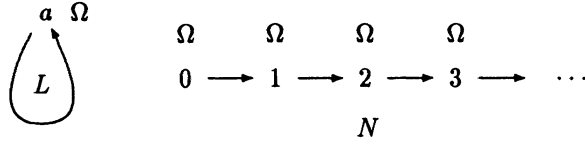


Figure B.3. Many pictures of $\Omega : (L, a)$ and each (N, n) .

xRz , so there exists some $y \in G_2$ such that xR_1y and yR_2z . If $u \leftarrow_1 x$, then by (B.31) for R_1 , there exists some $v \leftarrow_2 y$ such that uR_1v ; and then by (B.31) for R_2 , there exists some $w \leftarrow_3 z$ such that vR_2w , which together with uR_1v establish uRw . This is half of (B.31), and the other half is equally easy. $\dashv$

With these definitions, we can now prove that for a grounded graph G and any $p \in G$, the properties of (G, p) coded by the value $d_G(p)$ of its unique decoration are exactly those preserved under bisimulation of pointed graphs.

B.29. Theorem. *For all grounded graphs G and H with associated decorations d_G and d_H , and for all $p \in G$ and $q \in H$,*

$$d_G(p) = d_H(q) \iff (G, p) =_{\text{bs}} (H, q). \quad (\text{B.34})$$

Proof. We verify first that the relation

$$R = \{(x, y) \in G \times H \mid d_G(x) = d_H(y)\} \quad (\text{B.35})$$

satisfies (B.31), as follows:

$$\begin{aligned} xRy &\implies \{d_G(u) \mid u \leftarrow_G x\} = \{d_H(v) \mid v \leftarrow_H y\} \text{ by the def. of decoration} \\ &\implies (\forall u \leftarrow_G x)(\exists v \leftarrow_H y)[d_G(u) = d_H(v)] \\ &\quad \& (\forall v \leftarrow_H y)(\exists u \leftarrow_G x)[d_G(u) = d_H(v)] \\ &\implies (\forall u \leftarrow_G x)(\exists v \leftarrow_H y)uRv \\ &\quad \& (\forall v \leftarrow_H y)(\exists u \leftarrow_G x)uRv. \end{aligned}$$

Hence, if $p \in G$, $q \in H$, and $d_G(p) = d_H(q)$, then the relation R of (B.35) establishes that $(G, p) =_{\text{bs}} (H, q)$.

For the converse, suppose towards a contradiction that p is minimal in G such that there exists some $q \in H$ and a bisimulation R of the pointed graphs (G, p) and (H, q) , but $d_G(p) \neq d_H(q)$. Now

$$(\forall u \leftarrow_G p)(\exists v \leftarrow_H q)uRv,$$

hence,

$$(\forall u \leftarrow_G p)(\exists v \leftarrow_H q)[d_G(u) = d_H(v)]$$

by the choice of p , and, similarly, $(\forall v \leftarrow_H q)(\exists u \leftarrow_G p)[d_G(u) = d_H(v)]$, which proves $d_G(p) = d_H(q)$, contradicting the choice of p . $\neg$

It is a crucial property of **AFA** that it yields the same characterization of bisimulation for all graphs, by quite a different argument.

B.30. Theorem. (Aczel) (**AFA**) *For all graphs G and H with associated decorations d_G and d_H , and for all $p \in G$ and $q \in H$,*

$$d_G(p) = d_H(q) \iff (G, p) =_{\text{bs}} (H, q). \quad (\text{B.36})$$

Proof. The left-to-right implication in (B.36) is proved exactly as in **B.29**, that part of the argument did not depend on the given graphs being grounded.

Suppose now that the edge relations of G and H are $\rightarrow_G$ and $\rightarrow_H$ and $R \subseteq G \times H$ is a bisimulation of G with H . We can turn R into a pointed graph, with point the pair (p_G, p_H) and edge relation the product of $\rightarrow_G$ and $\rightarrow_H$:

$$(p, q) \rightarrow (u, v) \iff_{\text{df}} p \rightarrow_G u \ \& \ q \rightarrow_H v.$$

If d_G is the unique decoration of the graph G (forgetting the point) given by **AFA**, define on R the function

$$d_G^R(p, q) =_{\text{df}} d_G(p),$$

and compute:

$$\begin{aligned} x \in d_G^R(p, q) &\iff x \in d_G(p) \\ &\iff (\exists u \leftarrow_G p)[x = d_G(u)] \\ &\iff (\exists u \leftarrow_G p)(\exists v \leftarrow_H q)[u R v \ \& \ x = d_G(u)] \quad (\text{B.37}) \\ &\iff (\exists (u, v) \leftarrow (p, q))[x = d_G^R(u, v)], \end{aligned}$$

where the key equivalence (B.37) holds because R is a bisimulation and $p R q$, and hence for each $u \leftarrow_G p$, there exists some $v \leftarrow_H q$ satisfying $u R v$. Thus, the function d_G^R is a decoration of R , and the corresponding extension

$$d_H^R(p, q) =_{\text{df}} d_H(q)$$

of the decoration d_H of H is also a decoration of R , by the same argument. By **AFA** then, for all $(p, q) \in R$,

$$d_G(p) = d_G^R(p, q) = d_H^R(p, q) = d_H(q),$$

which completes the proof. $\neg$

This characterization under **AFA** of the properties of (G, p) which are coded into the value $d_G(p)$ of its unique decoration, suggests a method for the construction of $\mathcal{A}$.

B.31. The Antifounded Universe. Let

$$\mathcal{A}_0 =_{\text{df}} \{(G, \rightarrow_G, p_G) \in \mathcal{V} \mid \rightarrow_G \subseteq G \times G \ \& \ p_G \in G\} \quad (\text{B.38})$$

be the class of all pointed graphs on pure, grounded sets, and on $\mathcal{A}_0$ define the binary definite condition

$$(G, p_G) \varepsilon_0 (H, p_H) \iff_{\text{df}} (\exists q \in H)[q \leftarrow_H p_H \ \& \ (G, p_G) =_{\text{bs}} (H, q)], \quad (\text{B.39})$$

skipping the edge relations in the notation.

First we note that ε_0 respects bisimulation:

$$G_1 \varepsilon_0 H_1 \ \& \ G_1 =_{\text{bs}} G_2 \ \& \ H_1 =_{\text{bs}} H_2 \implies G_2 \varepsilon_0 H_2. \quad (\text{B.40})$$

To prove this, suppose $\rightarrow_1, p_1$ are the edge relation and the designated node of H_1 , and similarly with $\rightarrow_2, p_2$ for H_2 . The hypothesis of (B.40) gives us some $q_1 \leftarrow_1 p_1$ such that

$$G_2 =_{\text{bs}} G_1 =_{\text{bs}} (H_1, q_1), \quad (\text{B.41})$$

and a bisimulation R of (H_1, p_1) with (H_2, p_2) . By the basic property of bisimulations, there must exist some $q_2 \leftarrow_2 H_2$ such that $q_1 R q_2$; this means that R is a bisimulation of (H_1, q_1) with (H_2, q_2) , and then (B.41) with the transitivity of $=_{\text{bs}}$ gives $G_2 =_{\text{bs}} (H_2, q_2)$, hence $G_2 \varepsilon_0 H_2$.

Now each $G \in \mathcal{A}_0$ is a pure, grounded *set* (a triple in $\mathcal{V}$), even if it is ill founded *as a graph*, and the bisimulation condition $=_{\text{bs}}$ is an equivalence condition on $\mathcal{A}_0$ by **B.28**. By Problem **x12.22**, there exists a definite operation α which is determining for $=_{\text{bs}}$, i.e. for $G, H \in \mathcal{A}_0$

$$G =_{\text{bs}} H \iff \alpha(G) = \alpha(H).$$

The domain of the antifounded universe is the quotient class of $\mathcal{A}_0$ by $=_{\text{bs}}$,

$$\mathcal{A} =_{\text{df}} \{\alpha(G) \mid G \in \mathcal{A}_0\}. \quad (\text{B.42})$$

We define on $\mathcal{A}$ the membership condition

$$x \varepsilon y \iff_{\text{df}} (\exists G, H)[x = \alpha(G) \ \& \ y = \alpha(H) \ \& \ G \varepsilon_0 H], \quad (\text{B.43})$$

unambiguously by (B.40), and finally we take the *Pure Antifounded Universe* to be the triple $\mathcal{A}, \mathcal{A}, \varepsilon$. We will refer to it by the name of its domain $\mathcal{A}$, which is also the collection of its sets—there are no atoms in $\mathcal{A}$.

B.32. Theorem. (Aczel) **(AC)** $\mathcal{A}$ is a Rieger universe, which further satisfies the Antifoundation Principle **AFA**, the Axiom of Choice **AC** and the Principle of Purity.

Proof. The key property of $\mathcal{A}$ is that for each graph $H \in \mathcal{V}$ with edge relation $\rightarrow_H$ and each node $p \in H$,

$$\mathbf{b}_{\mathcal{A}}(\alpha(H, p)) = \{\alpha(H, q) \mid q \leftarrow_H p\}, \quad (\text{B.44})$$

which follows from the following trivial computation:

$$\begin{aligned} x \in \mathbf{b}_{\mathcal{A}}(\alpha(H, p)) &\iff (\exists G \in \mathcal{A}_0)(\exists q \leftarrow_H p)[x = \alpha(G) \ \& \ G =_{\text{bs}} (H, q)] \\ &\iff (\exists q \leftarrow_H p)[x = \alpha(H, q)]. \end{aligned}$$

This implies, in particular, that each $\mathbf{b}_{\mathcal{A}}(x)$ is a set, so $\mathcal{A}$ is a set universe.

For the Rieger property, suppose $Y \subseteq \mathcal{A}$ and (using **AC**) choose for each $y \in Y$ a pointed graph $G_y \in \mathcal{A}_0$, such that (1) $\alpha(G_y) = y$. By replacing each G_y by an isomorphic copy if necessary, we can also ensure that (2) $y \neq z \implies G_y \cap G_z = \emptyset$, and (3) for all $y \in Y$, $\emptyset \notin G_y$. Let

$$\begin{aligned} H &=_{\text{df}} \bigcup \{G_y \mid y \in Y\} \cup \{\emptyset\}, \\ u \rightarrow_H v &\iff_{\text{df}} (\exists y \in Y)[u \rightarrow_y v \vee (u = \emptyset \ \& \ v = p_y)], \end{aligned}$$

where $\rightarrow_y$ and p_y are the edge relation and the point of G_y . The pointed graph H with edge relation $\rightarrow_H$ and point $\emptyset$ is obviously in $\mathcal{A}_0$, and for each $y \in Y$,

$$(H, p_y) =_{\text{bs}} G_y \quad (\text{B.45})$$

by the trivial (identity) bisimulation

$$\{(u, v) \in H \times G_y \mid u = v\};$$

thus, by (B.44), and the definition,

$$\begin{aligned} \mathbf{b}_{\mathcal{A}}(\alpha(H, \emptyset)) &= \{\alpha(H, q) \mid q \leftarrow_H \emptyset\} \\ &= \{\alpha(H, p_y) \mid y \in Y\} \\ &= \{\alpha(G_y) \mid y \in Y\} \quad \text{by (B.45)} \\ &= Y. \end{aligned}$$

To prove the uniqueness of $\alpha(H)$, suppose H' is any pointed graph in $\mathcal{A}_0$ with edge relation $\rightarrow'$ and point q' such that

$$G \varepsilon_0 H' \iff \alpha(G) \in Y.$$

By (B.44) again,

$$\begin{aligned} \mathbf{b}_{\mathcal{A}}(\alpha(H', q')) &= \{\alpha(H', q) \mid q \leftarrow' q'\} \\ &= \{\alpha(H, p_y) \mid y \in Y\} \quad \text{by hyp.} \end{aligned}$$

Thus, for each $y \in Y$,

$$y = \alpha(G_y) = \alpha(H, p_y) \varepsilon \alpha(H'),$$

and we can choose (by **AC**) some $q_y \leftarrow' q'$ and a bisimulation S_y of (H, p_y) with (H', q_y) ; and conversely, by the same argument, for each $q \leftarrow' q'$ we can choose some $y_q \in Y$ and some bisimulation T_q of (H, q_y) with (H', q) . It is now easy to verify that the union

$$R = \bigcup \{S_y \mid p_y \leftarrow_H \emptyset\} \cup \{T_q \mid q \leftarrow' q'\} \cup \{(\emptyset, q')\}$$

is a bisimulation which establishes that $H =_{\text{bs}} H'$, i.e. $\alpha(H) = \alpha(H')$.

Finally, to verify **AFA** for $\mathcal{A}$, suppose G is a graph in $\mathcal{A}$ with edge relation $\rightarrow_G \in \mathcal{A}$. To prove that G admits a decoration in $\mathcal{A}$, it is enough to define an $\mathcal{A}$ -operation δ such that

$$\mathcal{A} \models (\forall p \in G)[\delta(p) = \{\delta(q) \mid q \leftarrow_G p\}], \quad (\text{B.46})$$

since $\mathcal{A}$ is a Z-F universe, so it “knows” from (B.46) that the restriction of δ to G is a function, which is then a decoration of G . Let

$$H =_{\text{df}} \mathbf{b}_{\mathcal{A}}(G) \in \mathcal{V},$$

and make H into a graph in $\mathcal{V}$ with the edge relation

$$x \rightarrow_H y \iff_{\text{df}} \mathcal{A} \models x \rightarrow_G y \quad (x, y \in H). \quad (\text{B.47})$$

For each $p \in H$, set

$$\delta(p) =_{\text{df}} \alpha(H, p) \in \mathcal{A} \quad (p \in H) \quad (\text{B.48})$$

and compute:

$$\begin{aligned} \mathbf{b}_{\mathcal{A}}(\delta(p)) &= \{\alpha(H, q) \mid q \leftarrow_H p\} && \text{by (B.44)} \\ &= \{\alpha(H, q) \mid \mathcal{A} \models q \leftarrow_G p\} && \text{by (B.47)} \\ &= \{\delta(q) \mid \mathcal{A} \models q \leftarrow_G p\} && \text{by (B.48)}. \end{aligned}$$

Put another way, for each $p \in G$,

$$x \in \delta(p) \iff \mathcal{A} \models (\exists q \leftarrow_G p)[x = \delta(q)],$$

which is equivalent to (B.46).

It remains to show that G admits at most one decoration in $\mathcal{A}$, and for this it suffices (as above) to show that if δ' is any $\mathcal{A}$ -operation such that

$$\mathcal{A} \models (\forall p \in G)[\delta'(p) = \{\delta'(q) \mid q \leftarrow_G p\}], \quad (\text{B.49})$$

then $\delta'(p) = \delta(p)$, for every $p \in H$. Given such a δ' , choose (by **AC**) a pointed graph H'_p with point r_p for each $p \in H$, such that

$$\delta'(p) = \alpha(H'_p) \quad (p \in H),$$

and make sure as in the proof of the Rieger property that these graphs are all pairwise disjoint. If $H' = \bigcup \{H'_p \mid p \in H\}$ is the union of all the graphs, then

$$\delta'(p) = \alpha(H', r_p) \quad (p \in H),$$

since (trivially) the identity relation $\{(q, q) \mid q \in H\}$ is a bisimulation of (H'_p, r_p) with (H, r_p) . We now claim that *the relation*

$$R =_{\text{df}} \{(p, s) \in H \times H' \mid \alpha(H', s) = \delta'(p)\}$$

is a bisimulation of (H, p) with (H', r_p) for each $p \in H$. This will complete the proof, because for $p \in H$, $\alpha(H', r_p) = \delta'(p)$, hence $p R r_p$, and hence

$$\delta(p) = \alpha(H, p) = \alpha(H', r_p) = \delta'(p).$$

To show the somewhat less trivial half of the italicized statement, let $\rightarrow'$ be the edge relation of H' , assume $p R s$ and compute:

$$\begin{aligned} t \leftarrow' s &\implies \alpha(H', t) \varepsilon \alpha(H', s) \\ &\implies \alpha(H', t) \varepsilon \delta'(p) && \text{because } p R s \\ &\implies \text{for some } q \leftarrow_H p, \alpha(H', t) = \delta'(q) && \text{by (B.49)} \\ &\implies \text{for some } q \leftarrow_H p, q R t. \end{aligned}$$

The Axiom of Choice for $\mathcal{A}$ follows from B.14, and the Principle of Purity is trivial. $\dashv$

Problems

xB.1. Prove that for each set universe $\mathcal{M} = M, S, E$, the axioms for definite conditions and operations listed in 3.18 become true, if we replace in them “condition” by “ $\mathcal{M}$ -condition”, “operation” by “ $\mathcal{M}$ -operation”, $\in$ by E , *Set* by S and $(\forall y)$ by $(\forall y \in M)$.

***xB.2.** Suppose pairs and Cartesian products are defined by the Kuratowski operation of 4.3. Show that $\bigcup_{n=2}^{\infty} \{N_0\}^n \notin \mathcal{Z}$ and infer that the following proposition is not a theorem of **ZDC**: *for each set A , there exists a function $f : N \rightarrow f[A]$ such that*

$$f(0) = A \times A, \quad f(n+1) = f(n) \times A.$$

***xB.3.** Construct a definite operation $(x, y)'$ with the following properties. (1) $(x, y)'$ is an ordered pair operation, i.e. it satisfies 4.1 and 4.2. (2) If $X, Y \in \mathcal{Z}$, then their Cartesian product $X \times Y$ is also in $\mathcal{Z}$. (3) If $\bigcup_{n=2}^{\infty} A^n$ is defined using this pair, then for each $A \in \mathcal{Z}$, $\bigcup_{n=2}^{\infty} A^n \in \mathcal{Z}$.

***xB.4.** Show that the implication **4.1** $\implies$ **4.2** is not a theorem of **ZDC** for an arbitrary definite operation (x, y) .

***xB.5.** Verify that if I is a transitive set, then

$$A \in M(I) \implies TC(A) \in M(I).$$

xB.6. For each I , define $K_n(I)$ by the recursion

$$K_0(I) = I, \quad K_{n+1}(I) = K_n(I) \cup \mathcal{P}(K_n(I)). \quad (\text{B.50})$$

Show that

$$M_n(I) \subseteq K_n(TC_{n+1}(I)) \subseteq M(I),$$

where $TC_n(I)$ and $M_n(I)$ are defined by (11.16) and (11.18), respectively.

***xB.7.** Find some $I \supset N_0$ such that $TC(I) \notin M(I)$. Infer that **ZDC** cannot prove that “every set has a transitive closure.”

***xB.8.** The implication (4.25) $\implies$ (4.27) cannot be proved for an arbitrary, definite operation $|A|$ in **ZDC**.

***xB.9.** The equivalence in Problem **x11.7** is not a theorem of **ZDC**.

***xB.10.** Assume that the full Axiom of Choice and the Generalized Continuum Hypothesis is true, so for all cardinals κ , $2^\kappa =_c \kappa^+$. Prove that

$$(n \mapsto \aleph_n) \subseteq \mathcal{Z}, \quad (n \mapsto \aleph_n) \notin \mathcal{Z}.$$

Infer that **ZAC** cannot prove the existence of an infinite, increasing sequence of infinite cardinals, i.e. the proposition

$$\theta : (\exists f : N \rightarrow f[N])(\forall n \in N)[N \leq_c f(n) <_c f(n+1)].$$

***xB.11.** Show that **ZDC** cannot prove the proposition “the wellordered set N of integers is similar with an ordinal”. **HINT:** Use Problem ***x12.8**. The less trivial part of the problem is how to compute (or avoid computing) the relativization of this fairly complex proposition.

***xB.12. (AC)** Show that **ZFC** cannot prove that strongly inaccessible cardinals exist. **HINT:** Go by contradiction and interpret the meaning of the alleged theorem in $\mathcal{V}_\kappa$, where κ is the least strongly inaccessible cardinal.

xB.13. An **ordered pair operation** in a Rieger universe $\mathcal{M}$ is any binary $\mathcal{M}$ -operation C such that for all $x, y, x', y' \in M$,

$$C(x, y) = C(x', y') \iff x = x' \ \& \ y = y'. \quad (\text{B.51})$$

Cartesian products and function spaces relative to C are defined by

$$\begin{aligned} A \times_C B &=_{\text{df}} \rho\{C(x, y) \mid xEA, yEB\}, \\ (A \rightarrow_C B) &=_{\text{df}} \rho\{f \in M \mid (\forall tEf)[tEA \times_C B] \\ &\quad \& (\forall xEA)(\exists!yEB)[C(x, y)Ef]\}, \end{aligned}$$

where $\rho(Y)$ is the Rieger operation of $\mathcal{M}$ defined in (B.14). Verify that these definitions make sense (i.e. ρ is applied to appropriate arguments) and hence $A \times_C B$ and $A \rightarrow_C B$ are $\mathcal{M}$ -operations.

***xB.14.** Define triples, structured sets and systems of natural numbers in an arbitrary Rieger universe $\mathcal{M}$, relative to an arbitrary ordered pair operation $C(x, y)$ in $\mathcal{M}$. Formulate the Choice Principles **DC**, **AC_N** and **AC** using these notions and prove that every Rieger universe $\mathcal{M}$ satisfies **DC**, and if **AC** is also true, then $\mathcal{M}$ also satisfies **AC**.

xB.15. Show that the Rieger universe $\mathcal{M}_a$ of **B.16** has an ordered pair operation C such that for all x and y , the “pair” $C(x, y)$ is an atom.

***xB.16. (AC)** Define a Rieger universe $\mathcal{M}$ which satisfies the following two propositions.

(a) There exists a binary, definite condition $\leq_a$ which well orders the class of atoms, in the sense that (1) for all atoms a, b, c ,

$$a \leq a, \quad [a \leq b \& b \leq c] \implies a \leq c, \quad [a \leq b \& b \leq a] \implies a = c,$$

(2) for every two atoms a, b , either $a \leq b$ or $b \leq a$, and (3) every non-empty set of atoms has a $\leq$ -least member.

(b) Every set X is equinumerous with an $\leq$ -initial segment of atoms, i.e. for some atom b , $X =_c \{a \mid \text{Atom}(a) \& a < b\}$.

HINT: Make the ordinals atoms in some Rieger universe.

***xB.17.** Define a Rieger universe which has at least two, distinct self-singletons, i.e. sets a and b such that $a \neq b$, $a = \{a\}$ and $b = \{b\}$. **HINT:** Start with a universe which has two atoms and imitate the coding construction in **B.16**.

***xB.18.** Define a Rieger universe which contains an infinite sequence of distinct sets $x_0, x_1, \dots$, such that for each i , $x_i = \{x_{i+1}\}$.

xB.19. Given a graph G and a node $p \in G$, let

$$G \restriction p = \{x \in G \mid x = p \vee p \Rightarrow x\}$$

consist of p and all the nodes on a path below it. Consider $G \upharpoonright p$ as a subgraph of G , with the restriction of the edge relation $\rightarrow_G$ to it, and prove that

$$(G, p) =_{\text{bs}} (G \upharpoonright p, p).$$

B.33. Definition. A **partial bisimulation** between two graphs G and H is any relation $R \subseteq G \times H$ which is a bisimulation of the pointed graphs (G, p) and (H, q) for every $(p, q) \in R$; it is a **total bisimulation** if in addition

$$(\forall p \in G)(\exists q \in H)p R q \ \& \ (\forall q \in H)(\exists p \in G)p R q.$$

Two graphs are **bisimilar** if there exists a total bisimulation between them.

xB.20. For all pairs of graphs G, H , there exists a largest (under $\subseteq$) partial bisimulation R between G and H , and $G =_{\text{bs}} H$ if and only if this largest bisimulation is total.

xB.21. Two graphs G, H are bisimilar if and only if every pointed graph (G, p) with $p \in G$ is bisimilar with some (H, q) , $q \in H$, and conversely, every (H, q) is bisimilar with some (G, p) .

xB.22. (AFA) Prove that there exists distinct, pure sets x, y and z such that

$$x \ni y \ni z \ni x,$$

and draw a picture of them.

***xB.23. (AFA)** Prove that there are only two, transitive, pure singletons. How many transitive, pure doubletons are there? Draw pictures of them.

xB.24. (AFA) With the Kuratowski pair, prove that there exists a pure set x such that

$$x = (\emptyset, x),$$

and draw a picture of it.

xB.25. (AFA) With the Kuratowski pair, prove that there exists a pure set x such that

$$x = \{(n, x) \mid n \in N\},$$

and draw a picture of it.

INDEX

- $(x \mapsto f(x))$, 3
- 0, 1, 2, 43
- $f : X \rightarrow Y$, 3
- $f : X \mapsto Y$, 3
- $f : X \rightarrow Y$, 3
- $f : X \mapsto Y$, 3
- $f(x) \downarrow, f(x) \uparrow$, 76
- $f : A \rightarrow E$, 76
- $f[A]$, 4
- $f \upharpoonright X$, 41
- N_F , 213
- $P =_o Q$, 96
- $U \subseteq V$, 95
- $[0, m)$, 64
- $[T]$, 150
- $\mathcal{A}$, 259
- $\aleph_0$, 58
- $\aleph_n$, 135
- $\aleph_\alpha$, 200
- $\approx$, 228
- $\beth_\alpha$, 205
- $=_{bs}$, 255
- $\perp$, 74
- $cf(\kappa)$, 141
- c , 69
- $\llbracket A/\sim \rrbracket$, 47
- $\chi(A), h(A)$, 106
- HF , 176, 239
- $u \mid v$, 150
- $\text{seg}_U(y)$, 95
- $\leq_c$, 8
- sup , 75
- Ω , 254
- $\theta^{(\mathcal{M})}$, 243
- $\langle x_n \rangle \rightsquigarrow (a, b)$, 223
- $=_c$, 7
- $\text{Succ}(P)$, 97
- $\mathcal{V}$, 183, 201
- $\mathcal{V}_\alpha$, 201
- $\mathcal{F}_\sigma, \mathcal{G}_\delta$, 153
- $\text{Function}(f)$, 41
- Absorption Laws, 138
- AC**, 117
- addition on N , 59
- AFA**, 254
- algebraic closure, 145
- algebraic numbers, 14
- analytic pointset, 153
- antifoundation, 254
- antifounded universe, 259
- archimedean, 221, 236
- asymptotic equivalence, 228
- atoms, 24, 30
- Axiom of Choice, 117, 120, 121, 125
 - equivalents, 120
- axiom of
 - Choice, **AC**, 117
 - Dependent Choices (**VI**), **DC**, 122
 - Extensionality (**I**), 24
 - Infinity (**VI**), 26
 - Pairset (**II**), 24
 - Powerset (**IV**), 25
 - Replacement (**VIII**), 170
 - Separation (subset) (**III**), 25
 - Unionset (**V**), 26

axiomatic setup, 23

axiomatic theory

ZDC, **ZAC**, 125

ZFC, 181

ZFDC, **ZFAC**, 170

axioms as closure properties, 30

axioms for Definite Conditions and
Operations, 27

axioms (I) - (VI), 24

Baire space, 147

Basic Closure Lemma, 175

Bekić-Scott rule, 114

best wellorderings, 135

bijection, 3

binary relation, 37

bisimulation, 255

partial, 265

Borel isomorphism, 166

Borel set, 160

AC_N, 122

Cantor set, 11

Cantor's Theorem, $A <_c \mathcal{P}(A)$,
15

Cantor-Bendixson Theorem, 151

cardinal number, 41, 199

cardinal

Assignment Problem, 42

by Frege, 206

by Scott, 208

by von Neumann, 198

Comparability Hypothesis, 19

Comparability, 121

Minimum Lemma, 138

regular, singular, 141

strongly inaccessible, 206

Supremum Lemma, 139

the next, 135

cardinality, 7

chain, 77

choice function, 119

choice set, 118

class, 28

closed pointset, 149

as body of a tree, 150

closed set, 45

cofinality, 141

Collapsing Lemma, 183

compact pointset, 155

Comparability of Well Ordered Sets,
104

complete ordered field, 221

composition, 4

congruence, 209

Consistency and Independence re-
sults, 126, 127, 164, 184

constructible sets, 142

constructively equivalent propo-
sitions, 127

Continuous Least Fixed Point The-
orem, 79

continuum, $\mathfrak{c}$, 69

Continuum Hypothesis, **CH**, 19

converse relation, 86

Countable Principle of Choice, **AC_N**,
122

countable set, 8

the rational integers, 10

the rationals, 10

countable unions of countable sets,
9

countably continuous, 79

cumulative rank hierarchy, 201

DC, 122

De Morgan's Laws, 5

decoration, Mostowski surjection,
183

Dedekind cuts, 229

Dedekind finite, infinite, 50, 131

definite conditions, operations, 20,
27

dense linear order, 219

Cantor characterization, 219

denumerable, 8

determining surjection, 210

directed-complete poset, 91

directed set, 91

disjoint union, 36

- enumeration, 8
- equinumerosity, 7
- equivalence class, 38
- equivalence condition, 207
 - Scott quotient, 207
- equivalence relation, 38, 209
 - determining surjection, 47
 - quotient, 47
- Euclidean algorithm, 84, 90
- expansive mapping, 98
- Extensionality Axiom, 24
- Fan Theorem, 134
- field, 211
 - complete ordered, 221
 - ordered, *see* ordered field, 212
- Finite Basis Lemma, 145
- finite cardinal, 64
- finite set, 8, 64
- finite set by Dedekind, 50
- Fixed Point Theorem, 108
 - detailed, 114
- Foundation (Regularity) Principle, 180
- Frege cardinals, 206
- function, 39
 - topologically continuous, 81
- General Comprehension Principle, 20
- Generalized Continuum Hypothesis, **GCH**, 19
- graph, 85
 - bisimulation, 255
 - grounded, 124
 - isomorphism, 255
 - pointed, 255
 - pointed picture, 255
- grounded $\in$ -recursion, 187
 - graph, 124
 - set, 179
- Hartogs' Theorem, 106
- Hausdorff space, 168
- hereditarily finite sets, **HF**, 176, 239
- hereditarily pure, finite, countable, 174
- ill founded set, 181
- image, 4
- independence of **CH**, 164
- indexed family, 40
- inductive poset, 77
 - directed complete, 145
- infinite by Dedekind, 50
- Infinity Axiom, 26
- initial segment, 95
- initial similarity, 103
- injection, 3
- Iteration Lemma, 101
- Kuratowski pair, 34
- König's Lemma, 133
- König's Theorem, 140
- Least Fixed Point Theorem, 108
 - Continuous, 79
- least Zermelo universe, $\mathcal{Z}$, 179
- linear ordering, 61
- linearization, 71, 144
- Liouville's Theorem, 14
- mapping, 39
 - continuous, 80, 145
 - countably continuous, 79
 - expansive, 98
 - iteration, 101
 - monotone, 78
 - order-preserving, 96
- marriage, 72
- Maximal Chain Principle, 121
- minimal point, 87
- model, 239, 243
- Model Existence Results, proviso, 127
- models of axiomatic theories, 126
- monotone mapping, 78
- Mostowski Collapsing Lemma, 183
- Mostowski surjection, decoration, 183
- multiplication on N , 59

- natural numbers, 53, 58
 - existence, 54
 - order, 62
 - uniqueness, 54
- Nested Interval Property, 226
- normal ordinal operation, 205
- open covering, 168
- open pointset, 149
- open set, 45
- orbit, 79
 - under an operation, 173
- order-preserving function, 96
- ordered field, 212
 - archimedean, 221
 - complete, 221
 - non-archimedean, 236
- ordered pair operation, 35
- ordinal, 190
 - addition and multiplication, 197
 - assignment characterization, 193
 - characterization, 194
 - comparison, 195
 - exponentiation, 204
 - normal operation, 205
 - order, 196
 - properties first, 191
 - properties second, 192
 - properties third, 192
 - recursion, 197
 - successor, 196
- Pairset Axiom, 24
- partial function, 76
 - finite, 80
- partial ordering, 61
- partially ordered set, poset, 73
- Peano axioms, 53
- perfect pointset, 150
 - cardinality, 150
- Perfect Set Theorem, 157
- picture, 255
- Pigeonhole Principle, 64
- pointed graph, 255
- pointset, 148
 - $\mathcal{F}_\sigma, \mathcal{G}_\delta$, 152
 - analytic, 153
 - Borel, 160
 - closed, 149
 - compact, 155
 - open, 149
 - perfect, 150
- pointwise ordering, 86
- poset, 73
 - directed-complete (dcpo), 91
 - discrete, 74
 - flat, 74
 - inductive, 77
 - inductive and dcpo, 145
 - product, 110
 - successor, 97
 - sum, 110
- powerset, 14
- Powerset Axiom, 25
- prewellordering, 112
- Principle of
 - Antifoundation, 254
 - Foundation, 180
 - General Comprehension, 20
 - Purity, 30
 - Soundness of Logical Inference, 244
- property **P**, 152
- proposition, 243
- proviso, for model existence results, 127
- purity, 30
- rational numbers
 - countable, 10
 - existence, 215
 - uniqueness, 214
- real numbers, 221, 236
 - completeness, 13, 221
 - existence, 231
 - uncountable, 11
 - uniqueness, 234

- recursion, 55
 - grounded into $\mathcal{W}$, 172
 - grounded long, 188
 - grounded on $\in$, 187
 - on N , 55
 - on N complete, 72
 - on N into $\mathcal{W}$, 173
 - on N simultaneous, 66
 - on N with parameters, 59
 - on ON , 197
 - transfinite, 100
 - transfinite into $\mathcal{W}$, 173
- regular cardinal, 141
- Regularity (Foundation) Principle, 180
- relation, 37
- relativization, 243
- Replacement Axiom, 170
- restriction, $f \upharpoonright X$, 41
- Rieger universe, 248
- Rieger's Theorem, 248
- Russell paradox, 21
- Schröder-Bernstein Theorem, 16
 - Zermelo's proof, 50
- Scott topology, 91
- self singleton, 181, 254
- Separation Axiom, 25
- sequence
 - bounded, 224
 - Cauchy, 223
 - converging, 223
 - settling, 223
- set, 1
 - Borel, 160
 - closed, 45
 - grounded, 179
 - ill founded, 181
 - open, 45
 - partially ordered, 73
 - pure, grounded, 202
 - structured, 45
 - universe, 242
 - well orderable, 94
 - well ordered, 93
- similarity, $=_o$, 96
 - initial, 103
- Soundness of Logical Inference, 244
- space, 45
- splitting tree, 150
- stream, 87
- string, 67
- strongly inaccessible cardinal, 206
- structured set, 45
- Subset Axiom, 25
- successor poset, $Succ(P)$, 97
- surjection, 3
- system of natural numbers, 53
- ternary relation, 39
- The Next Cardinal κ^+ , 135
- theory
 - ZDC**, **ZAC**, 125
 - ZFC**, 181
 - ZFDC**, **ZFAC**, 170
- topological space, 46
- topology, 45
 - about, 81
 - of pointwise convergence, 90
 - of Scott, 91
- Transfinite Induction Theorem, 98
- Transfinite Recursion Theorem, 100
- transformation, 39
- transitive closure, 174
- transitive set, class, 174
- tree, 132
 - finitely branching, 133
- uncountable reals, 11
- uncountable set, 8
 - of binary sequences, 11
- Unionset Axiom, 26
- universe
 - of sets, 242
 - antifounded, 259
 - natural, 243
 - Rieger, 248
 - Z-F, 183
 - Zermelo, 177
- vector space basis, 145

von Neumann cardinals, 198
 von Neumann surjection, $\mathbf{v}$, 190
 von Neumann's class $\mathcal{V}$, 183

 well founded graph, 124
 well orderable set, 94
 well ordered set, 93
 well ordered set comparability, 104
 Wellfoundedness of $\leq_c$, 134
 Wellfoundedness of $\leq_o$, 105
 wellordering, 62
 Wellordering Theorem, 121

Wiener pair, 46

ZDC, **ZFC**, 125
 Zermelo Fraenkel Set Theory, **ZFC**,
 181
 Zermelo universe, 177
 least, $\mathcal{Z}$, 179
 Zermelo-Fraenkel (Z-F) universe,
 183
ZFC, 181
ZFDC, **ZFAC**, 170
 Zorn's Lemma, 121

Undergraduate Texts in Mathematics

(continued)

Lidl/Pilz: Applied Abstract Algebra.

Macki-Strauss: Introduction to Optimal Control Theory.

Malitz: Introduction to Mathematical Logic.

Marsden/Weinstein: Calculus I, II, III. Second edition.

Martin: The Foundations of Geometry and the Non-Euclidean Plane.

Martin: Transformation Geometry: An Introduction to Symmetry.

Millman/Parker: Geometry: A Metric approach with Models. Second edition.

Moschovakis: Notes on Set Theory.

Owen: A First Course in the Mathematical Foundations of Thermodynamics.

Palka: An Introduction to Complex Function Theory.

Pedrick: A First Course in Analysis.

Peressini/Sullivan/Uhl: The Mathematics of Nonlinear Programming.

Priestley: Calculus: An Historical Approach.

Protter/Morrey: A First Course in Real Analysis. Second edition.

Protter/Morrey: Intermediate Calculus. Second edition.

Ross: Elementary Analysis: The Theory of Calculus.

Samuel: Projective Geometry.

Readings in Mathematics.

Scharlau/Opolka: From Fermat to Minkowski.

Sigler: Algebra.

Silverman/Tate: Rational Points on Elliptic Curves.

Simmonds: A Brief on Tensor Analysis. Second edition.

Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.

Smith: Linear Algebra. Second edition.

Smith: Primer of Modern Analysis. Second edition.

Stanton/White: Constructive Combinatorics.

Stillwell: Mathematics and Its History.

Strayer: Linear Programming and Its Applications.

Thorpe: Elementary Topics in Differential Geometry.

Troutman: Variational Calculus with Elementary Convexity.

Valenza: Linear Algebra: An Introduction to Abstract Mathematics.